

Université de Montréal

**L'évolution de l'édition des ARN de transfert
chez les jakobides**

par

Jessica Leigh

Département de Biochimie

Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en biochimie

août, 2003

© Jessica Leigh, 2003



W

4

U58

2004

v.007

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

**L'évolution de l'édition des ARN de transfert
chez les jakobides**

présenté par :

Jessica Leigh

a été évalué par un jury composé des personnes suivantes :

Président-rapporteur: Serguei Chteinberg

Directeur de recherche: Franz Lang

Membre du jury: Léa Brakier-Gingras

Mémoire accepté le : _____

RESUME

Les jakobides sont des protistes bactérivores avec le génome mitochondrial le plus ancestral retrouvé chez les eucaryotes, dans le sens qu'il ressemble à un génome α -protéobactérien. A cause du caractère bactérien du génome mitochondrial des jakobides, on s'attendait à ce que les mécanismes d'expression des gènes soient également bactériens. Cependant, plusieurs mésappariements ont été trouvés dans la région de la tige acceptrice de deux gènes d'ARNt lors du séquençage du génome mitochondrial du jakobide *Seculamonas ecuadoriensis*. Nous avons séquencé les ADNc pour ces deux ARNt par RT-PCR pour investiguer cette observation. Nos résultats indiquent que les trois dernières positions de l'extrémité 3', plus le nucléotide discriminateur, des ARNt glutamyl et séryl sont changés suite à la transcription, ce qui produit des paires de bases de type Watson-Crick entre ces positions, et ce qui remplace le nucléotide discriminateur par une adénosine. Ce processus d'édition ressemble à celui identifié chez le métazoaire *Lithobius forficatus*. Chez *S. ecuadoriensis*, le nucléotide situé le plus en amont de la séquence éditée n'est pas toujours lui-même édité, ce qui indique un mécanisme d'édition qui commence par la dégradation exonucléolytique de l'extrémité 3', suivie par sa resynthèse. Ce mécanisme d'édition 3' des ARNt est probablement distinct de celui identifié chez *L. forficatus*, malgré les similarités entre les deux systèmes. De plus, ces résultats suggèrent que la maturation des ARNt mitochondriaux chez *S. ecuadoriensis* suit un mécanisme qui ressemble à celui retrouvé chez *Escherichia coli*, au lieu des mécanismes observés dans les organites des autres eucaryotes.

Mots clés : Edition des ARN; Jakobides; Maturation des ARNt; Réparation des ARNt; Evolution de la mitochondrie.

ABSTRACT

The jakobid flagellates are bacterivorous protists with mitochondrial genomes that are the most ancestral identified to date, in that they most resemble the genomes of the α -proteobacterial ancestors of the mitochondrion. Because of the bacterial character of jakobid mitochondrial genomes, it was expected that mechanisms for gene expression and RNA structures would be bacterial in nature. However, sequencing of the mitochondrial genome of the jakobid *Seculamonas ecuadoriensis* revealed several apparent mismatches in the acceptor stems of two tRNA genes. To investigate this observation, we determined the cDNA sequences of these tRNAs by RT-PCR. Our results show that the last three positions of the 3' extremity plus the discriminator position of seryl and glutamyl tRNAs are altered post-transcriptionally, restoring orthodox base pairing and replacing the discriminator with an adenosine residue, in an editing process that resembles that of the metazoan *Lithobius forficatus*. However, the furthest upstream of the edited nucleotides is occasionally left unedited, indicating that the editing mechanism proceeds initially by exonucleolytic degradation, followed by repair of the degraded region. This 3' tRNA editing mechanism is likely distinct from that of *L. forficatus*, despite the apparent similarities between the two systems. Furthermore, these results strongly suggest the presence of a tRNA 3' processing mechanism in jakobid mitochondria that resembles that observed in *Escherichia coli*, rather than the processing mechanisms described in the organelles of other eukaryotes.

Key Words: RNA editing; Jakobids; tRNA processing; tRNA repair; Mitochondrial evolution.

TABLE DES MATIERES

Titres	Pages
RESUME.....	III
ABSTRACT	IV
TABLE DES MATIERES	V
LISTE DES TABLEAUX	VIII
LISTE DES FIGURES	IX
LISTE DES CODES DES ACIDES AMINES	XII
REMERCIEMENTS	XIII
CHAPITRE 1 : INTRODUCTION	1
1.1 Structure et maturation des ARNt	2
1.1.1 Structure des ARNt.....	2
1.1.2 Maturation de l'extrémité 3' des ARNt chez les eubactéries	4
1.1.3 Maturation de l'extrémité 3' des ARNt nucléaires chez les eucaryotes.....	5
1.1.4 Maturation de l'extrémité 3' des ARNt mitochondriaux et chloroplastiques	6
1.2 L'édition des ARN	7
1.2.1 L'édition d'ARNm dans les kinétoplastides.....	7
1.2.2 Les deux mécanismes d'édition des ARN mitochondriaux chez <i>Physarum polycephalum</i>.....	8
1.2.3 Édition co-transcriptionnelle chez les paramyxovirus	10
1.2.4 L'édition des ARNm nucléaires chez les animaux	11
1.2.5 L'édition par amination et désamination dans les organites des plantes	11
1.2.6 L'édition en 5' des ARNt mitochondriaux chez <i>Acanthamoeba</i> et chez les chytridiomycètes.....	12

1.2.7	L'édition des ARNt mitochondriaux chez les métazoaires	12
1.3	Caractéristiques structurales, classification, et évolution des jakobides	15
1.3.1	Caractéristiques structurales des jakobides	15
1.3.2	Description de l'ADN mitochondrial des jakobides	16
1.3.3	Taxonomie et phylogénie des jakobides et des eucaryotes	20
CHAPITRE 2 : ARTICLE		22
2.1	Abstract	24
2.2	Introduction	25
2.3	Materials and Methods	28
2.3.1	Cell culture, cloning, sequencing, and annotation of mtDNA	28
2.3.2	Phylogenetic analysis	28
2.3.3	RNA purification	28
2.3.4	Oligonucleotides	29
2.3.5	RT-PCR, cloning, and sequencing.....	29
2.4	Results and Discussion	30
2.4.1	Phylogenetic analysis indicates that <i>S. ecuadoriensis</i> and <i>R. americana</i> are monophyletic.....	30
2.4.2	Editing of mitochondrial seryl and glutamyl tRNAs in <i>S. ecuadoriensis</i>	30
2.4.3	3' tRNA editing suggests a eubacterial mechanism for tRNA processing.....	35
2.4.4	<i>S. ecuadoriensis</i> mitochondrial 3' tRNA editing analogous to <i>L. forficatus</i> ?	40
2.5	Acknowledgements	41
2.6	References	42

CHAPITRE 3 : DISCUSSION ET CONCLUSION	43
3.1 Discussion	44
3.1.1 Édition des ARNt chez <i>Seculamonas ecuadoriensis</i>	44
3.1.2 Évolution récente du mécanisme d'édition	45
3.1.3 Modèle théorique pour le mécanisme d'édition	45
3.1.4 Implications pour la maturation de l'extrémité 3' des ARNt.....	46
3.1.5 Modèle théorique pour l'évolution de l'édition des ARNt chez <i>S. ecuadoriensis</i>	47
3.2 Conclusion	48
BIBLIOGRAPHIE.....	49

LISTE DES TABLEAUX

<u>Tableaux</u>	<u>Pages</u>
ARTICLE	
Tableau II.1	Non-Watson-Crick base pairs in aminoacyl acceptor stems32
Tableau II.2	Summary of tRNA sequencing results37

LISTE DES FIGURES

<u>Figures</u>	<u>Pages</u>
INTRODUCTION	
Figure I.1	Structure secondaire des ARNt3
Figure I.2	Mécanisme d'édition par insertion et délétion de U chez les kinétoplastides.....9
Figure I.3	Images des jakobides 17
Figure I.4	Taille des génomes mitochondriaux et ensemble de gènes mitochondriaux 18
Figure I.5	Cartes des ADNmt des jakobides 19
Figure I.6	Phylogénie globale des eucaryotes21
ARTICLE	
Figure II.1	Global eukaryotic phylogeny31
Figure II.2	Inferred tRNA secondary structures.....34
Figure II.3	Processing, editing, and repair of tRNAs36

LISTE DES ABREVIATIONS

A	Adénosine
ADAR	Adénosine désaminase qui agit sur l'ARN
ADN	Acide déoxyribonucléique
ADNc	ADN complémentaire
ADNk	ADN kinétoplastidique
ADNmt	ADN mitochondrial
ApoB	Apolipoprotéine B
ARN	Acide ribonucléique
ARNg	ARN guide
ARNm	ARN messenger
ARNt	ARN de transfert
ARNr	ARN ribosomique
ARNtm	ARN de transfert/messenger
<i>atp</i>	Gènes des sous-unités de la ATP synthetase
ATP	Adénosine triphosphate
C	Cytosine
CCase	ARNt nucléotidyltransférase
<i>cob</i>	Gène de la apocytochrome B
<i>cox</i>	Gènes des sous unités de la cytochrome oxidase
<i>cox2</i>	Gène de la sous-unité 2 de la cytochrome oxidase
G	Guanine
I	Inosine
kpb	Kilo paires de bases
ML	Maximum de vraisemblance
<i>nad</i>	Gènes des sous-unités de la NADH déshydrogénase
PNPase	Polynucléotide phosphorylase
RNase	Ribonucléase
RnRp	ARN polymérase ARN-dépendante
<i>rpl</i>	Gènes des protéines de la grande sous-unité du ribosome
<i>rpo</i>	Gènes des sous-unités de la ARN polymérase

<i>rps</i>	Gènes des protéines de la petite sous-unité du ribosome
<i>sdh</i>	Gènes des sous-unités de la succinate déshydrogénase
RT-PCR	Reverse transcription PCR
T	Thymine
<i>tufA</i>	Gène du facteur d'élongation Tu
U	Uracile

LISTE DES CODES DES ACIDES AMINES

A	Ala	Alanine
B	Asx	Asparagine ou acide aspartique
C	Cys	Cystéine
D	Asp	Acide aspartique
E	Glu	Acide glutamique
F	Phe	Phénylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Méthionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Sérine
T	Thr	Thréonine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine
Z	Glx	Glutamine ou acide glutamique

REMERCIEMENTS

J'aimerais d'abord remercier mon directeur de recherche, le Dr Franz Lang, pour le soutien et la disponibilité qu'il m'a offerts tout au long de ce projet, et pour le temps qu'il a investi à la rédaction des articles et à ce mémoire de maîtrise.

Je dois ma reconnaissance sincère à Lise Forget pour la lecture critique de ce mémoire, et de tout ce que j'ai écrit en français durant les deux dernières années. Je la remercie aussi pour les conseils techniques qu'elle m'a offerts, et pour sa compagnie, grâce à laquelle ma maîtrise a été beaucoup plus agréable.

Je remercie Isabelle Nadeau, qui m'a motivée de venir à Montréal et de poursuivre une maîtrise dans une université francophone. Je la remercie aussi pour les nombreuses corrections de français qu'elle a faites pour moi, et pour tout le soutien et l'encouragement qu'elle m'a offerts depuis que je l'ai connue.

Je veux remercier tous les membres du laboratoire du Dr F. Lang, surtout Elias Seif et Dennis Lavrov, pour leurs nombreux conseils, et Zhang Wang, pour son assistance technique.

Finalement, j'aimerais remercier les membres du laboratoire du Dr Gertraud Burger pour leur assistance technique et pour l'accès à l'équipement, et surtout Shona Teijeiro, pour ses photographies des jakobides.

CHAPITRE 1 : INTRODUCTION

L'édition d'un ARN est définie par la modification de sa séquence primaire pour produire une séquence qui ne correspond plus à celle du gène duquel il a été transcrit, mais qui aurait pu être codée par la séquence génomique (1). Ce terme s'applique à divers processus enzymatiques, mais n'inclus pas, par exemple, les modifications qui produisent les bases non-standards des ARNt. L'édition des ARN a été identifiée chez un grand nombre d'espèces, principalement chez les eucaryotes, dans les organites. L'identification d'un mécanisme d'édition de l'extrémité 3' des ARN de transfert (ARNt) mitochondriaux chez *Seculamonas ecuadoriensis*, un membre des jakobides, notre attente étant de trouver des mécanismes d'expression des gènes mitochondriaux plutôt bactérien, est surprenant; par contre, nous croyons que ce résultat indique un mécanisme de maturation en 3' des ARNt qui ressemble à celui des bactéries. L'introduction de ce mémoire est une revue de la littérature sur la maturation et l'édition des ARN et sur les jakobides. Aussi sont incluses des données non-publiées concernant le génome mitochondrial de *S. ecuadoriensis*.

1.1 Structure et maturation des ARNt

1.1.1 Structure des ARNt

Les ARNt, qui jouent un rôle central dans l'expression des gènes, servent de molécules adaptatrices entre les ARNm et les protéines pour lesquelles ils codent. Ils sont caractérisés par leur structure secondaire en forme de « feuille de trèfle » (Figure I.1a). Ces molécules se replient pour former quatre tiges : l'acceptrice, la T, la D, et l'anticodon. Ces trois dernières molécules se terminent en boucle, tandis que la tige acceptrice est formée par les extrémités 3' et 5' de la molécule. Dans un ARNt mature, l'extrémité 3' dépasse la 5' par quatre nucléotides incluant le nucléotide discriminatoire, suivi par la séquence 5'-CCA-3'. Durant l'aminocyclation de l'ARNt, le discriminatoire sert d'élément identifiant de l'ARNt à l'ARNt aminosynthétase; l'acide aminé est ensuite attaché à l'adénosine terminale.

Figure I.1 : Structure secondaire des ARNt

a) Structure secondaire de l'ARNt^{Ala} de *E. coli*. Cette structure secondaire a été repliée en se basant sur la séquence primaire du gène retrouvé dans le génome d'*E. coli* K12 (Numéro d'accèsion Genbank NC_000913). Cette structure est un exemple de structure typique d'un ARNt, avec une tige acceptrice de sept paires de bases, une tige T de cinq paires de bases, une tige anticodon de cinq paires de bases, et une tige D de quatre paires de bases. b) Structures secondaires d'ARNt mitochondriaux chez les métazoaires *Mytilus californianus* (2) et *Lithobius forficatus* (3). Dans ces structures, la longueur des tiges est plus variable. Dans certains cas, les tiges T et D sont même manquantes.

tRNA^{Ala}
(*E. coli*)

B

mt-tRNA^{Met}
(*L. forficatus*)

mt-tRNA^{His}
(*L. forficatus*)

mt-tRNA^{Ser}
(M. californianus)

La conservation de la structure secondaire des ARNt est presque universelle. Cependant, chez les métazoaires, la structure des ARNt mitochondriaux est plus variable. On retrouve souvent des molécules avec des tiges tronquées ou allongées. La figure I.1b présente quelques unes de ces structures.

La maturation des ARNt inclut plusieurs étapes, y compris la modification de certaines bases, le clivage de la séquence de tête à l'extrémité 5' et de la séquence de queue à l'extrémité 3', et, dans certains cas, l'édition et/ou l'épissage. Malgré que la maturation de l'extrémité 5' soit accomplie par un seul clivage endonucléolytique, catalysé par la RNase P, le mécanisme par lequel la séquence en queue est enlevée est plus variable.

1.1.2 Maturation de l'extrémité 3' des ARNt chez les eubactéries

Chez les bactéries, les études sur la maturation de l'extrémité 3' des ARNt ont été faites majoritairement chez *Escherichia coli*. Conséquemment, le modèle qui décrit la maturation de l'extrémité 3' des ARNt chez les bactéries tient compte de plusieurs aspects du génome d'*E. coli* (position des ARNt entre d'autres gènes, CCA terminal codé par le génome, longues extensions en 5' et en 3' des précurseurs ARNt).

Chez *E. coli*, la maturation commence probablement par un clivage endonucléolytique situé plusieurs bases en aval de l'extrémité 3'; cette étape n'est possiblement pas nécessaire dans tous les cas (4). L'extrémité est ensuite dégradée par plusieurs exonucléases. Durant ce processus, la RNase P, qui est plus active quand il y a peu de séquence d'excès à l'extrémité 3', catalyse la coupure endonucléolytique de la séquence de tête. La dégradation de l'extrémité 3' se continue suite à l'action de la RNase P, et se termine peu après. L'ordre de ces événements est variable, selon les concentrations relatives des enzymes.

Il y a six exonucléases qui agissent durant la maturation de l'extrémité 3' des ARNt chez *E. coli* : la RNase II, la RNase D, la RNase BN, la RNase T, la RNase

PH, et la polynucléotide phosphorylase (PNPase). Il semble qu'au moins cinq de ces enzymes agissent *in vivo* (toutes sauf la PNPase, dont le rôle dans la maturation des ARNt n'est pas facile à démontrer, dû à sa participation dans plusieurs processus essentiels), et que leurs rôles dans la maturation de l'extrémité 3' se chevauchent. Cependant, la RNase II et la PNPase sont les plus efficaces sur les ARNt avec une très longue extension 3' (>10 nucléotides) et les RNases PH et T sont généralement les plus efficaces sur des ARNt à courte extension (<5 nucléotides; 4-7).

Une autre enzyme qui agit dans la voie de la maturation en 3' des ARNt est l'ARNt nucléotidyltransférase (CCase), l'enzyme responsable de la synthèse matrice-indépendante de la séquence terminale CCA. Chez *E. coli*, cette séquence est codée par le génome; conséquemment, la CCase n'est pas essentielle, car son rôle n'est pas dans la synthèse du CCA, mais dans sa réparation (8). La maturation de l'extrémité devient alors une compétition entre la dégradation et la resynthèse, qui se termine seulement quand l'extrémité est protégée contre la dégradation par son aminoacylation.

1.1.3 Maturation de l'extrémité 3' des ARNt nucléaires chez les eucaryotes

Les ARNt nucléaires chez les eucaryotes sont transcrits par l'ARN polymérase III, et leurs extensions 3' et 5' sont très courtes (5-10 et 6-15 nucléotides respectivement). Généralement, la maturation de l'extrémité 3' des ARNt nucléaires, qui se passe dans le noyau, est le résultat d'un seul clivage endonucléolytique. Chez la levure *Saccharomyces cerevisiae*, l'étude *in vitro* détaillée de Papadimitriou et Gross (9) a démontré que la maturation de l'extrémité 3' est catalysée par des exonucléases et par des endonucléases. Cependant, il a été établi que, *in vivo*, seules les endonucléases y participent normalement; l'action de ces enzymes nécessite la participation de la protéine Lhp1p, qui stabilise probablement la conformation de l'ARNt. Le rôle des exonucléases semble être de remplacer les endonucléases si la Lhp1p est absente (10).

La maturation de l'extrémité 3' chez les plantes est catalysée par la RNase Z, qui clive la molécule immédiatement en aval du discriminateur. Cette endonucléase peut agir sur un ARNt partiel qui contient au moins la tige T et la tige acceptrice, mais elle est beaucoup moins efficace que si la molécule est intacte. La RNase Z est un monomère, et elle a une enzyme homologue chez les archaea qui clive aussi l'extrémité 3' suite à la maturation de l'extrémité 5'. Il existe aussi un enzyme homologue à la RNase Z chez les eubactéries, où son rôle est essentiel, malgré qu'il soit inconnu (11-15).

1.1.4 Maturation de l'extrémité 3' des ARNt mitochondriaux et chloroplastiques

Chez la plupart des eucaryotes, la maturation de l'extrémité 3' des ARNt mitochondriaux est effectuée par un seul clivage endonucléolytique à la position du discriminateur. Ceci est le cas chez la levure, les plantes, et les animaux (13, 16, 17).

Dans le cas des mitochondries animales, le génome est tellement compact que les ARNt sont généralement directement adjacents, ou peuvent même se chevaucher. Comme le génome mitochondrial (ADNmt) est transcrit en entier, il n'est pas surprenant que les ARNt soient séparés par un clivage endonucléolytique : il n'y a pas assez de séquence en excès entre les ARNt pour permettre l'action d'une exonucléase. Cependant, il existe sans doute une activité exonucléolytique qui peut agir sur l'extrémité 3' des ARNt mitochondriaux, quoique ce ne soit pas nécessairement son rôle principal dans la mitochondrie (18).

Dans les chloroplastes, les ARNt sont séparés par un clivage endonucléolytique qui laisse un nucléotide de plus après le discriminateur. Ce nucléotide, codé par le génome, est généralement une cytidine. Conséquemment, la CCase n'a qu'à rajouter un CA pour que la maturation de l'extrémité 3' soit complète (19).

1.2 L'édition des ARN

1.2.1 L'édition d'ARNm dans les kinétoplastides

Le groupe taxonomique Kinetoplastida, comprenant les Trypanosomatidae et Bodonidae, inclut plusieurs parasites. Parmi les kinétoplastides pathogènes, on retrouve *Trypanosoma brucei*, l'agent cause de la trypanosomiase d'Afrique (la maladie du sommeil), et *Leishmania mexicana*, qui cause des lésions cutanées chez les patients infectés. Les maladies causées par les kinétoplastides sont contractées à travers un vecteur d'infection, généralement un insecte qui se nourrit de sang humain.

Les kinétoplastides sont caractérisés par la présence d'une seule, énorme mitochondrie qui se trouve à la base du flagelle, dans laquelle se trouve le kinétoplaste, la structure qui contient l'ADN kinétoplastidique (ADNk). L'ADNk est un réseau d'environ 50 maxicercles et 10 000 minicercles caténés. Leurs gènes structuraux sont codés par les maxicercles, tandis que les minicercles codent pour les ARN guides (ARNg).

Le séquençage du gène mitochondrial *cox2*, qui code pour la sous-unité 2 de la cytochrome oxidase, chez *T. brucei* a indiqué la présence d'un déphasage multiple du cadre de lecture. Cependant, il ne semblait pas être un pseudogène, car sa séquence était très conservée parmi les trypanosomes. Le séquençage de l'ARNm a résolu le mystère : il contenait quatre nucléotides U qui n'étaient pas codés par le génome mitochondrial (20). Ce fut le premier cas identifié de l'édition d'un ARN.

L'édition des ARNm par insertion ou, plus rarement, délétion d'uridines est prévalent chez les kinétoplastides, mais n'a jamais été observée ailleurs. Dans certains cas, l'édition des ARNm est tellement extensive que leurs gènes ne peuvent être identifiés. Par exemple, chez *T. brucei*, les gènes qui codent pour les sous-unités 7 et 9 de la NADH déshydrogénase produisent des transcrits primaire auxquels 551 et 345 U y sont respectivement ajoutés. Il existe également des transcrits qui ne sont pas édités, ou qui sont très peu édités (par exemple, *cox2*).

Le mécanisme d'édition chez les kinétoplastides est guidé par les ARNg. Ces petits ARN contiennent une courte séquence antisens qui peut s'apparier au transcrit primaire en aval du site d'édition. Le transcrit est ensuite modifié par l'ajout ou la délétion d'uridines pour que sa séquence soit complémentaire à celle de l'ARNg (Figure I.2). Plusieurs ARNg sont nécessaires pour l'édition d'un seul transcrit; l'édition guidée par un ARNg produit le site d'ancrage pour le prochain, ce qui mène à une chaîne d'évènements d'édition dans la direction de 3' à 5'.

1.2.2 Les deux mécanismes d'édition des ARN mitochondriaux chez *Physarum polycephalum*

Bien que l'édition des ARN chez les kinétoplastides soit l'exemple le plus extensif, l'édition des ARN mitochondriaux chez le myxomycète *Physarum polycephalum* est probablement l'exemple le plus compliqué. L'édition chez *P. polycephalum* emploie l'insertion des quatre nucléotides, A, C, G, et U, et aussi de dinucléotides. De plus, la conversion de nucléotides C en U a été observée. Ce dernier mécanisme est moins rapide que le premier, et il est probablement le résultat d'une désamination. L'insertion des nucléotides est co-transcriptionnelle, ce qui indique que l'ARN polymérase pourrait participer à l'édition.

Le mécanisme employé par *Physarum* pour guider l'édition de ses ARN n'est pas encore connu, mais il n'utilise pas les ARNg, comme chez kinétoplastides. De plus, il ne peut pas utiliser un mécanisme d'insertion comme celui des kinétoplastides, où une séquence d'ancrage sur le côté 3' du site d'édition permet l'attachement d'une molécule guide, car l'édition est trop rapide chez *P. polycephalum* pour avoir une assez longue séquence en aval du site d'édition. Le mécanisme d'édition par insertion est très efficace, et il affecte les ARNm, les ARNr, et les ARNt. Il a également été identifié chez quatre autres myxomycètes, *Stemonitis flavogenita*, *Didymium nigripes*, *Arcyria cinerea*, et *Clastoderma debaryanum*, où l'édition est plus extensive chez les deux premiers (1, 21-23).

Figure I.2 : Mécanisme d'édition par insertion et délétion de U chez les kinétoplastides

a) L'édition commence par l'hybridation entre l'ARNg et le pré-ARNm; b) l'insertion (et la délétion) d'uridines produit une séquence complémentaire à celle de l'ARNg, et produit un site d'ancrage pour le deuxième ARNg; c) et d) le processus continue avec un deuxième ARNg. Après plusieurs rondes d'édition, un ARNm mature est produit qui peut servir à la traduction(24).

1.2.3 Édition co-transcriptionnelle chez les paramyxovirus

Les paramyxovirus utilisent l'édition des ARNm pour produire deux cadres de lecture à partir du gène P. Cette famille de virus, qui inclut les virus de la rougeole, du para-influenza, du Sendai, et de la rubéole, a un génome ARN simple brin dont un ARNm monocistronique est transcrit. L'antigénome, un intermédiaire de la réplication, est transcrit du génome aussi. Les paramyxovirus emploient un mécanisme d'édition qui insère jusqu'à dix G dans l'ARNm à un site spécifique. Ces nucléotides sont incorporés lors de la transcription.

L'insertion des G est le résultat de la présence dans le génome d'une région riche en pyrimidines composée d'une longue série de U suivie de plusieurs C. Lors de la transcription, l'ARN polymérase virale ralentit, et quand elle atteint le début des C, elle s'arrête, insérant deux G (ou plus) pour le premier C sur le brin complémentaire. Chez le virus de l'Ebola, un mécanisme semblable insère des résidus A.

Un résultat intéressant fut la découverte de l'insertion des G lors de la transcription de l'antigénome. Malgré que, *in vivo*, ce phénomène soit limité à la transcription de l'ARNm, une modification de la longueur du génome peut causer l'insertion de nucléotides G à cette étape. Normalement, le nombre de nucléotides dans le génome des paramyxovirus est un multiple de six, ce qui est nécessaire pour la formation des nucléocapsides, structures qui servent comme matrice pour la transcription, et qui sont des complexes nucléoprotéiniques qui contiennent six nucléotides. Dans les expériences *in vitro* de Hausmann et ses collègues (25), avec un génome contenant un nombre de nucléotides non multiple de six, cette structure hexamère était rétablie dans l'antigénome, et conséquemment les copies futures du génome, par l'insertion de nucléotides G au même site que lors de la transcription de l'ARNm.

1.2.4 L'édition des ARNm nucléaires chez les animaux

L'édition des ARN nucléaires est rare chez les animaux. Les deux exemples qui seront présentés ici sont des cas de l'édition par désamination. La lipoprotéine de basse densité (LDL) transporte le cholestérol du foie aux tissus périphériques. Chaque particule LDL contient une seule apolipoprotéine B (apoB). L'ARNm de l'apoB est édité par une désamination C à U, ce qui produit un codon non-sens UAA. Le résultat est la production de deux protéines, l'apoB-100 et l'apoB-48, à partir d'un seul gène. Les deux formes de l'apoB sont exprimées dans différents tissus; chez l'humain, l'apoB-100 est exprimée dans le foie, tandis que l'apoB-48 est exprimée dans l'intestin grêle.

Les désaminations A à I sont utilisées dans l'édition de plusieurs ARNm nucléaires chez les animaux. Le nucléotide I s'apparie au C lors de la traduction, donc cette réaction a le même effet qu'une substitution d'un A par un G. La famille d'enzymes responsable de la plupart des événements de déamination A à I est la famille ADAR (les adénosine désaminases qui agissent sur l'ARN). L'ADAR2, qui est responsable de l'édition de plusieurs ARNm chez les mammifères, est elle-même éditée. Le résultat de la désamination d'un nucléotide A dans l'ARNm de l'ADAR2 chez le rat est la production d'un site accepteur d'épissage 3' alternatif qui a pour effet de produire quatre protéines à partir d'un seul gène (26).

1.2.5 L'édition par amination et désamination dans les organites des plantes

Dans les organites de la plupart des plantes terrestres qui ont été étudiées jusqu'à présent, les ARN subissent de nombreuses conversions de C à U et de U à C; les aminations (U à C) sont beaucoup plus rares que les désaminations. Les événements d'édition chez les plantes sont beaucoup plus communs dans la mitochondrie, où jusqu'à 1000 sites de désamination peuvent exister, que dans les chloroplastes. Les ARNm, les ARNt, et, plus rarement, les ARNr sont tous affectés par l'édition.

1.2.6 L'édition en 5' des ARNt mitochondriaux chez *Acanthamoeba* et chez les chytridiomycètes

Le premier exemple d'édition des ARNt a été démontré chez *Acanthamoeba castellanii*, un amoebozoaire (27). L'ADNmt de cette espèce code pour quinze ARNt, dont douze contiennent au moins un mésappariement parmi les trois premières paires de bases de la tige acceptrice. Ces douze ARNt sont édités aux trois premières positions de l'extrémité 5' pour produire des paires de bases de type Watson-Crick (27, 28). L'activité enzymatique responsable de l'édition des ARNt mitochondriaux chez *A. castellanii* a été partiellement purifiée, et les résultats d'expériences *in vitro* indiquent que l'édition est le résultat de l'action d'une exonucléase ou d'une endonucléase, suivi par une nucléotidyltransférase 3'-à-5', qui agit d'une façon matrice-dépendente (29).

Un mécanisme d'édition analogue a été identifié chez les chytridiomycètes, un groupe de champignons. Certains membres de ce groupe ont un système d'édition des ARNt, tandis que d'autres en sont dépourvus. La distribution phylogénétique de l'édition de l'extrémité 5' des ARNt indique que ce processus est apparu indépendamment au moins trois fois au cours de l'évolution : une fois chez *Acanthamoeba*, et deux fois chez les chytridiomycètes (30, 31, M-J Laforest et BF Lang, données non-publiées).

1.2.7 L'édition des ARNt mitochondriaux chez les métazoaires

Dans plusieurs cas chez les animaux, les ARNt codés par l'ADNmt ressemble à ceux que l'on trouve chez *A. castellanii*, avec une tige acceptrice qui contient des mésappariements. Ceci est souvent causé par la taille minuscule de l'ADNmt animal, dans lequel les ARNt partagent fréquemment un ou plusieurs nucléotides. Dans le cas du calmar *Loligo bleekeri*, les deux G qui semblent coder pour les deux dernières positions de l'extrémité 3' de l'ARNt^{Tyr} sont partagés avec son voisin, l'ARNt^{Cys}, où

ils codent pour les deux premières positions de l'extrémité 5'. Mais dans l'ARNt^{Tyr} mature, ces deux G sont remplacés par deux A (32). Ceci représente un cas d'édition de l'extrémité 3' de la tige acceptrice (tandis que, chez *A. castellanii*, l'extrémité 5' est éditée.)

Un cas semblable existe chez le poulet *Gallus gallus*, où l'ARNt^{Tyr} partage son discriminateur, un G, avec l'ARNt^{Cys}. Encore une fois, ce G est remplacé par un A dans l'ARNt^{Tyr}. Dans la plupart des ARNt^{Tyr} séquencés, le discriminateur est suivi par le CCA qui caractérise un ARNt mature, mais dans certains, les auteurs ont trouvé une queue de 2 à 13 nucléotides A à la place du CCA. Pour cette raison, les auteurs proposent qu'un clivage endonucléolytique produit une extrémité 5' mature chez l'ADNt^{Cys} et une extrémité 3' sans discriminateur chez l'ARNt^{Tyr}. La poly(A)polymérase ajouterait ensuite une queue poly(A) à l'ARNt^{Tyr}. Finalement, un deuxième clivage endonucléolytique enlève les nucléotides A en excès, et la CCase termine la maturation de l'ARNt (33).

Il paraît qu'un mécanisme semblable est présent chez l'escargot *Euhadra herklotsi*, où l'extrémité 3' des ARNt mitochondriaux chevauche parfois jusqu'à six nucléotides du gène adjacent. Dans cette espèce-ci, une grande partie de la tige acceptrice est réparée par une activité enzymatique qui semble appartenir à la poly(A)polymérase, car les nucléotides insérés sont toujours des A (34).

Chez l'ornithorynque *Ornithorhynchus anatinus*, il est moins probable que l'édition des ARNt mitochondriaux soit due à l'action de la poly(A)polymérase. Dans l'ADNmt de cette espèce, l'extrémité 3' de l'ARNt^{Ser} chevauche l'ARNt^{Leu} par quatre nucléotides; deux de ces quatre nucléotides sont mésappariés (formant des paires G:A et A:C), et un troisième forme une paire de base G:U dans l'ARNt^{Ser}. Les auteurs ont démontré que l'identité de deux de ces nucléotides, plus celle du discriminateur, sont changées au niveau de l'ARN (l'identité du C qui fait partie du mésappariement A:C n'est pas changée). Comme l'insertion de nucléotides C est impliquée dans l'édition de cet ARNt, il est peu probable qu'elle soit catalysée par la poly(A)polymérase (35).

Il existe un mécanisme de réparation des ARNt mitochondriaux chez l'humain qui ressemble beaucoup à l'édition identifiée ailleurs. L'ADNmt humain, comme celui de plusieurs autres métazoaires, contient des ARNt qui chevauchent l'extrémité 5' du gène adjacent. Chez l'humain, l'identité des nucléotides qui forment l'extrémité 3' de la tige acceptrice de l'ADNt n'est pas changée. Cependant, l'équipe de Reichert (36) a démontré que, dans le cas de l'ARNt^{Tyr}, qui partage son discriminateur, un A, avec l'ARNt^{Cys} adjacent, un clivage endonucléolytique le laisse sans discriminateur; l'ARNt^{Cys} demeure intact. Le discriminateur de l'ARNt^{Tyr} est rajouté par une enzyme qui ressemble beaucoup à la CCase en terme d'activité enzymatique : quoiqu'elle rajoute toujours un A *in vivo*, un C est parfois inséré *in vitro*. Comme les deux nucléotides qui précèdent le discriminateur sont des C, les auteurs suggèrent que l'on retrouve une légère modification à l'activité réparatrice qui existe normalement chez la CCase, ce qui lui laisse réparer le discriminateur.

Cependant, l'enzyme qui rajoute le discriminateur à l'ARNt^{Tyr} mitochondrial humain ne répare pas nécessairement que le discriminateur. Des essais *in vitro* ont démontré que cette activité enzymatique peut aussi réparer des ARNt auxquels ils manquent jusqu'à quatre nucléotides. Ces résultats ont mené à l'hypothèse du modèle dynamique de la réparation des ARNt. Ce modèle explique l'édition et la réparation matrice-indépendante des ARNt mitochondriaux chez les métazoaires. Il propose que, comme l'extrémité 3' des ARNt est souvent incomplète (phénomène causé soit par la dégradation exonucléolytique, soit par le clivage endonucléolytique qui le sépare du gène adjacent), elle soit complétée par une forme dérivée de la CCase, qui insère des nucléotides de façon aléatoire. Si les nucléotides forment des paires de bases Watson-Crick avec l'extrémité 5', la structure est reconnue par l'aminoacyl-ARNt-synthétase. Sinon, l'ARN est susceptible à la dégradation exonucléolytique, et la réparation recommence, jusqu'à ce que l'ARNt soit aminoacylé (18).

L'édition des ARNt mitochondriaux chez le centipède *Lithobius forficatus* est la plus extensive chez les animaux. Tous sauf un des vingt-deux ARNt codés par l'ADNmt contiennent des mésappariements dans la région de la tige acceptrice. Malgré que certains des gènes des ARNt se chevauchent, ce n'est pas toujours le cas.

La séquence de huit ARNt mitochondriaux a indiqué que, dans chacun de ces cas, l'extrémité 3' de l'ARNt est éditée par le remplacement des nucléotides mésappariés. Le nombre élevé de molécules éditées indique que ce mécanisme est présent dans l'ADNmt de *L. forficatus* depuis longtemps. Mais en plus du grand nombre d'ARNt édités, *L. forficatus* utilise une combinaison des quatre nucléotides pour éditer ses ARNt. Pour cette raison, les auteurs suggèrent que l'édition des ARNt mitochondriaux chez *L. forficatus* est guidée par l'extrémité 5', qui agit comme matrice pour une ARN polymérase ARN-dépendante, une enzyme possiblement héritée d'un virus (3). Schürer et ses collègues (12) suggèrent que ce n'est qu'un autre cas qui suit le modèle dynamique de la réparation des ARNt de Reichert et Mörl (18), qui implique la synthèse matrice indépendante de l'extrémité 3', mais que l'enzyme qui la synthétise peut accepter n'importe quel nucléotide.

Contrairement à la règle où l'édition des ARNt mitochondriaux affecte l'extrémité 3' chez les métazoaires, une exception a été identifiée. Chez les marsupiaux, l'ADNmt contient un ARNt avec l'anticodon qui spécifie la glycine (GCC) à l'endroit où les autres mammifères ont l'ARNt^{Asp}(GUC). Il a été démontré que, dans 50% des cas, cet anticodon est édité. De plus, les deux formes de l'ARNt sont fonctionnelles, et sont aminoacylées avec l'acide aminé correspondant (glycine quand l'anticodon demeure GCC, et aspartate quand l'anticodon devient GUC; 37).

1.3 Caractéristiques structurales, classification, et évolution des jakobides

1.3.1 Caractéristiques structurales des jakobides

Les jakobides sont des protistes biflagellés qui se nourrissent de bactéries. Ces organismes unicellulaires mesurent moins de 10 µm de longueur et sont caractérisés par une rainure qui traverse en longueur du côté ventral de la cellule, qu'ils utilisent pour l'ingestion des bactéries. Quand elle se nourrit, la cellule s'attache à sa proie avec le flagelle antérieur et crée un courant d'eau vers la rainure ventrale avec le flagelle postérieur, ce qui dirige les bactéries vers ce point d'ingestion (38-40).

La figure I.3 présente des images des jakobides *Jakoba bahamiensis*, *Jakoba libera*, *Reclinomonas americana*, et *Seculamonas ecuadoriensis*.

1.3.2 Description de l'ADN mitochondrial des jakobides

La mitochondrie, l'organite qui produit l'énergie pour les cellules eucaryotiques, a comme ancêtre un endosymbiote qui faisait partie de la sous-division alpha des protéobactéries. L'érosion du génome de l'endosymbiote a laissé un ADNmt qui code pour un nombre de gènes très limité. La mitochondrie chez les jakobides est la plus primitive, en termes du caractère bactérien de son génome, parmi toutes les mitochondries étudiées jusqu'à présent.

L'ADNmt chez les jakobides contient un nombre de gènes sans précédent. Comme l'indique la Figure I.4, l'ADNmt de *R. americana* contient un superensemble des gènes trouvés dans toute autre mitochondrie. Le complément génique chez les autres jakobides ressemble beaucoup à celui de *R. americana* (Figure I.5).

En plus de son complément génique très élevé, l'ADNmt chez les jakobides contient aussi plusieurs éléments qui indiquent la présence de mécanismes eubactériens pour l'expression des gènes. L'ADNmt de *R. americana* contient un gène qui code pour la RNase P (le ribozyme qui est responsable de la maturation de l'extrémité 5' des ARNt) qui contient tous les éléments du consensus minimal bactérien pour cette molécule, et aussi un ARNtm dérivé (41, 42). Il code aussi pour les quatre sous-unités d'une ARN polymérase de type bactérien; chez tout autre eucaryote, cette enzyme est remplacée par une enzyme nucléaire qui semble provenir d'un bactériophage. Le facteur d'élongation Tu est aussi codé par l'ADNmt de *R. americana*. Finalement, l'ordre des gènes de l'ADNmt chez les jakobides est à la fois logique et bien conservé, ce qui indique la présence d'unités transcriptionnelles qui ressemblent aux opérons bactériens (42; voir aussi la Figure I.5). Aucun de ces éléments typiquement eubactériens n'a été observé dans les ADNmt hors des jakobides.

Figure I.3 : Images des jakobides

a) *Jakoba bahamiensis*; b) et c) *J. libera*; d) et e) *R.. americana*; f) et g) *S. ecuadoriensis*. Sur ces images photographiques, on peut observer les deux flagelles qui sont typiques des jakobides.

A



B



C



D



E



F



G



Figure I.4 : Taille des génomes mitochondriaux et ensemble de gènes mitochondriaux

a) Les génomes circulaires et linéaires sont respectivement représentés par un cercle et une ligne. Pour les génomes de taille supérieure à 60 kpb, l'ADN qui code pour les gènes de fonction connue (rouge) est distingué de l'ADN non-codant et celui qui code pour des cadres de lecture (ORF) non-identifiés (bleu). Les noms des espèces sont *C. (Chlamydomonas) reinhardtii*, *C. (Chlamydomonas) eugametos*, et *S. (Schizosaccharomyces) pombe*. Le génome de *Rickettsia prowazekii* (une α -protéobactérie) est représenté pour comparaison. Les jakobides sont représentés par les génomes de *J. libera* et *R. americana*. b) Chaque ellipse contient les gènes mitochondriaux présents chez l'organisme qu'elle représente. Seuls les gènes qui codent pour des protéines et pour les ARNr sont représentés. L'ensemble de gènes mitochondriaux présent chez le jakobide *R. americana* contient tous les gènes mitochondriaux identifiés chez tout autre organisme (43).

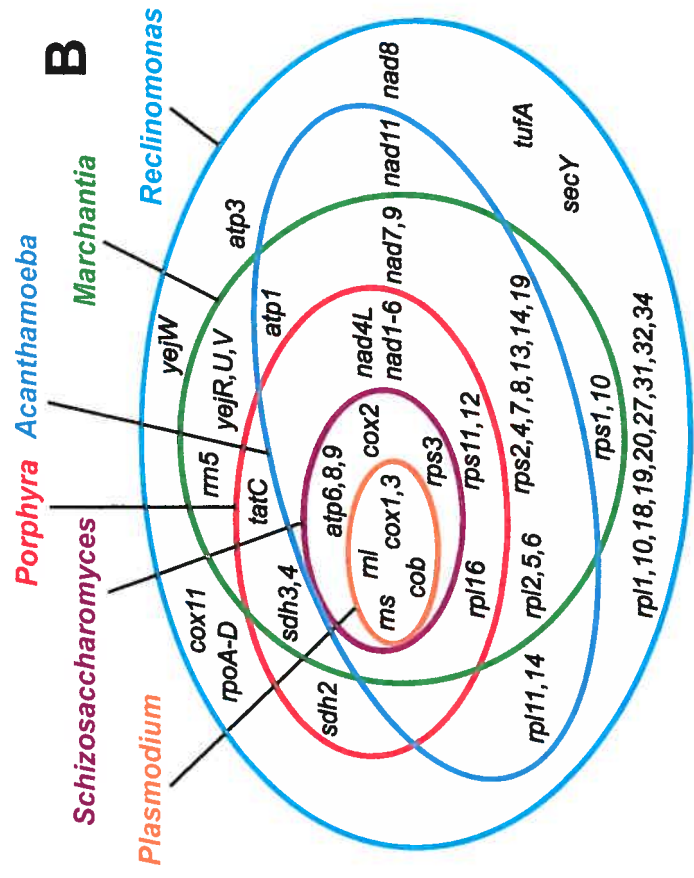
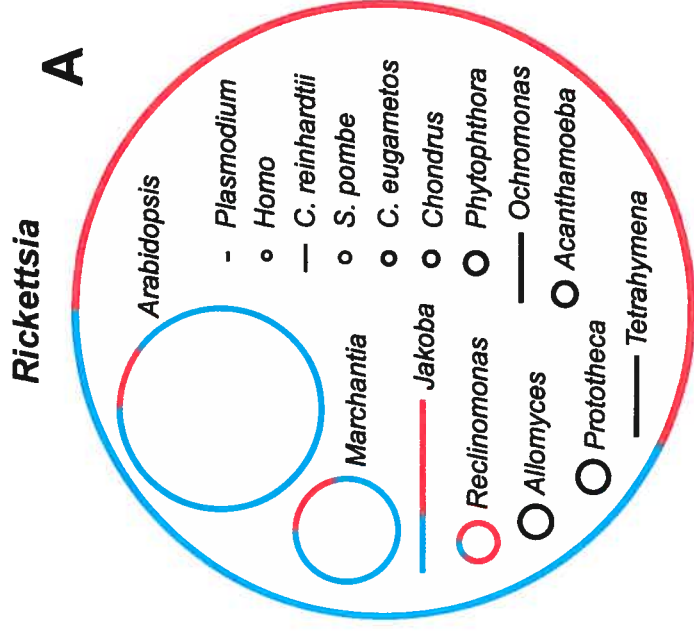
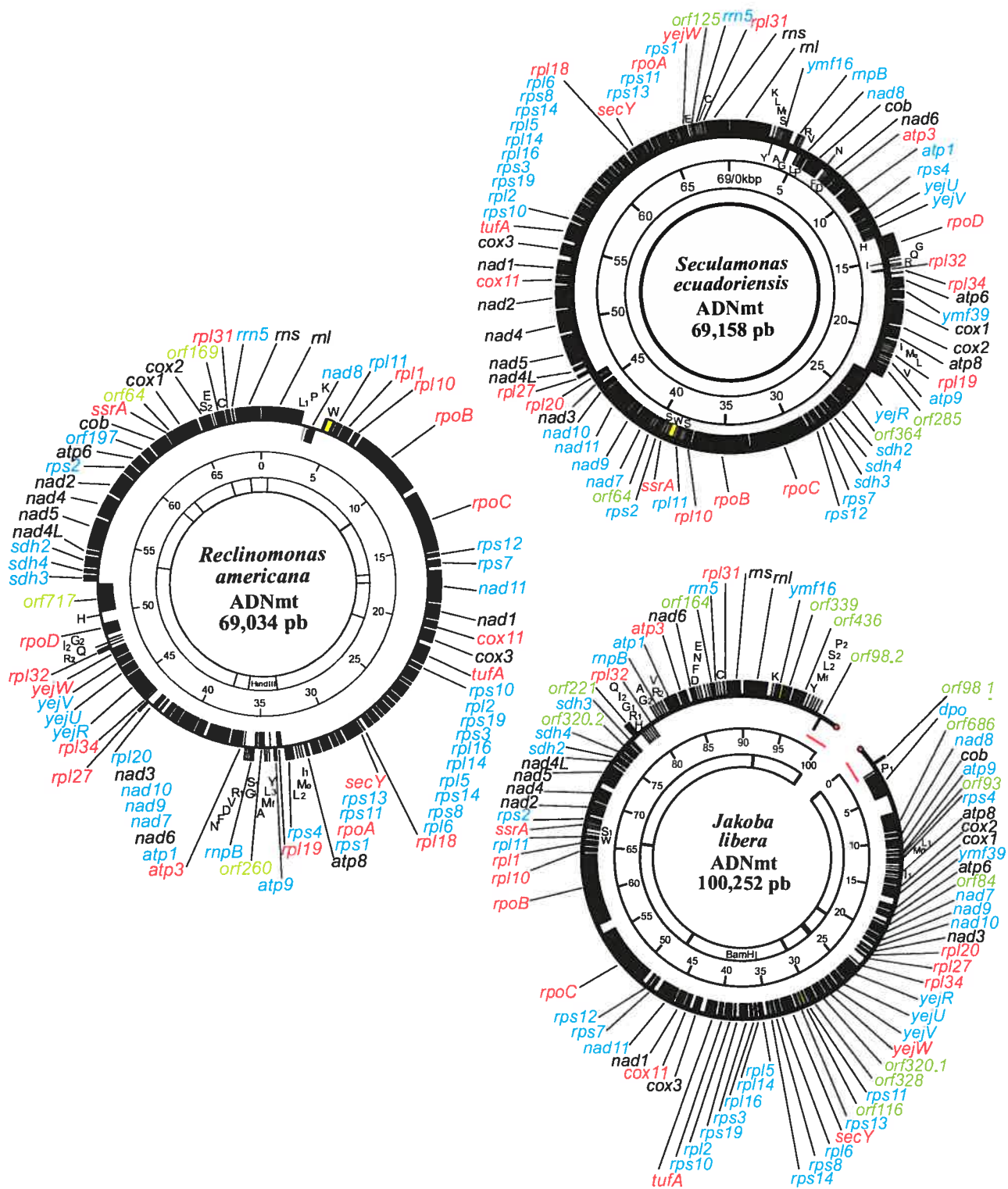


Figure I.5 : Cartes des ADNmt des jakobides

La carte génique de l'ADNmt est présentée pour les jakobides *R. americana* (42), *S. ecuadoriensis*, et *J. libera*. Les gènes qui codent pour des ARNt sont identifiés par les codes une-lettre des acides aminés. Les introns sont représentés par des rectangles jaunes, et les gènes, par des rectangles noirs. Les gènes qui sont à l'extérieur du cercle sont transcrits de gauche à droite, tandis que ceux qui sont à l'intérieur sont transcrits de droite à gauche. Rouge, gènes qui sont trouvés uniquement chez les jakobides; bleu, gènes qui sont absents chez les vertébrés, mais qui sont souvent présents chez les plantes et chez les protistes; vert, ORFs qui sont trouvés uniquement chez les jakobides.



1.3.3 Taxonomie et phylogénie des jakobides et des eucaryotes

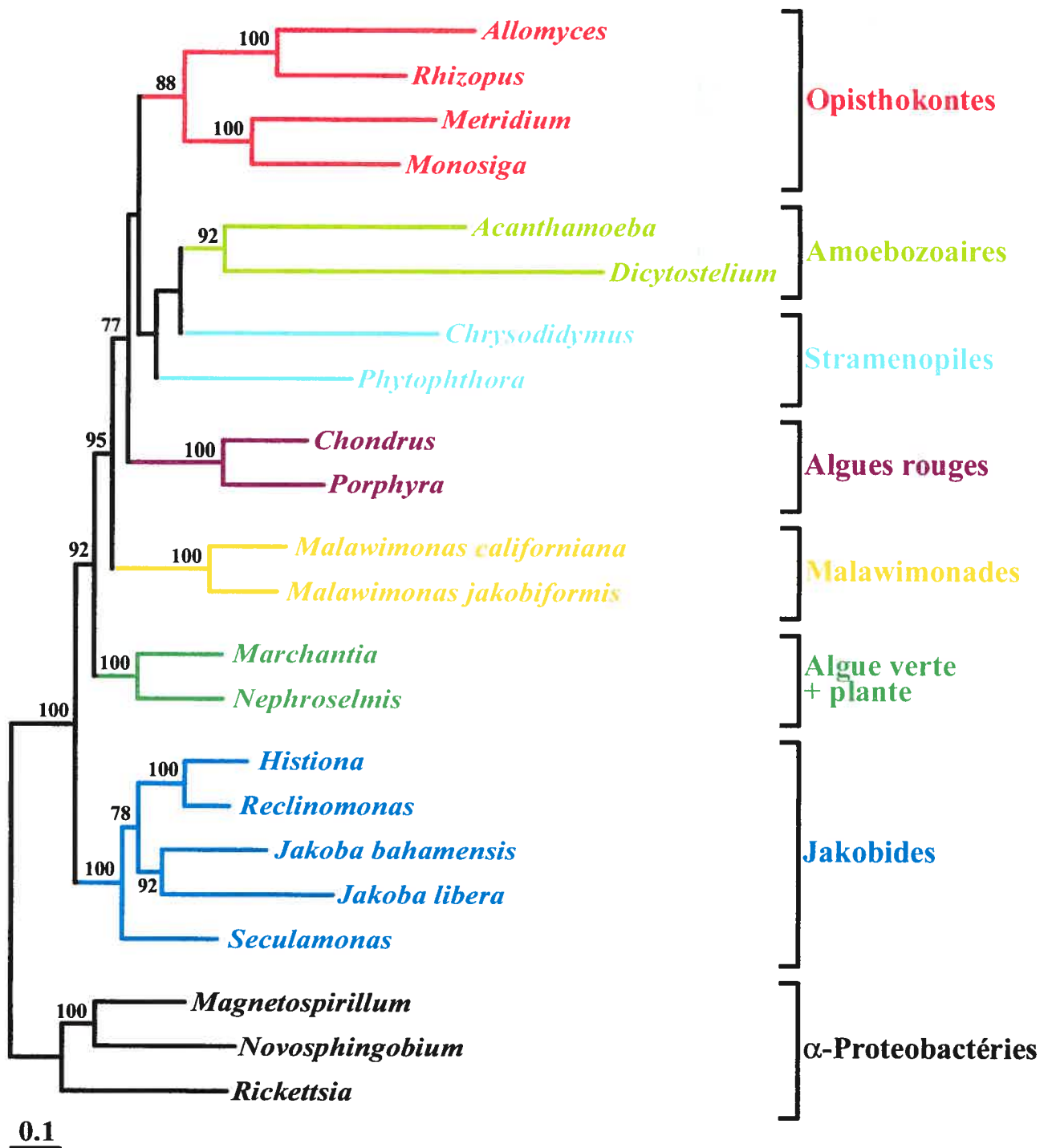
Les données morphologiques ont indiqué une affiliation entre les familles Malawimonadidae (ou malawimonades) et Jakobidae (ou jakobides). Cependant, les données moléculaires démontrent que les deux familles sont distantes au niveau phylogénétique (Figure I.6). Le nom « jakobides » est donc utilisé uniquement pour décrire les membres de la famille Jakobidae (qui inclut les genres *Reclinomonas*, *Jakoba*, *Histiona*, et *Seculamonas*.)

La position phylogénétique des jakobides dans l'arbre eucaryotique n'est pas bien résolue. Ce groupe est souvent regroupé avec les plantes et les algues vertes, mais ceci est probablement le résultat d'une attraction entre les courtes branches de ces deux groupes, et n'est que faiblement supporté. En même temps, la position des jakobides est mieux résolue quand des données provenant de plusieurs jakobides sont utilisées. La figure I.6 indique que les jakobides forment le groupe le plus ancestral chez les eucaryotes, avec une valeur bootstrap de 92%; cet arbre démontre aussi la monophylie des jakobides, avec une valeur bootstrap de 100%. Cependant, la topologie exacte à l'intérieur de ce groupe n'est pas bien résolue, mais il semble que *S. ecuadoriensis* se branche à la base du groupe.

Figure I.6 : Phylogénie globale des eucaryotes

Cet arbre a été reconstruit à partir des séquences protéiques des gènes *atp*, *cob*, *cox*, *nad*, *rpl*, *rpo*, *rps*, *sdh*, et *tufA*, codés par l'ADNmt. La reconstruction a été faite par maximum de vraisemblance (ML) avec le logiciel ProML avec une valeur alpha de 0.85, calculée par CodeML (44). Les valeurs de bootstrap ML ont été calculés basé sur 100 générations, produit par Seqboot et l'arbre consensus a été produit par Consense. ProML, Seqboot, et Consense font partie du progiciel Phylip (Felsenstein, J., 2002; v. 3.6a 2.1, distribué par l'auteur).

Le regroupement des jakobides est très bien supporté par un bootstrap de 100%. La position de *S. ecuadoriensis* à la base des jakobides est faiblement supportée par un bootstrap de 78%, tandis que le regroupement de *J. libera* et de *J. bahamensis* est mieux supporté par un bootstrap de 92%, et le regroupement de *Histiona aroides* et de *R. americana* est supporté par un bootstrap de 100%. La position des jakobides à la base des eucaryotes est supportée par un bootstrap de 92%, ce qui indique un support moyen. Les séquences ont été trouvées chez Genbank : *Allomyces macrogynus* (NC_001715), *Metridium senile* (NC_000933), *Monosiga brevicollis* (NC_004309), *Acanthamoeba castellanii* (NC_001637), *Dictyostelium discoideum* (NC_000895), *Chrysodidymus synuroideus* (NC_002174), *Phytophthora infestans* (NC_002387), *Chondrus crispus* (NC_001677), *Porphyra purpurea* (NC_002007), *Malawimonas jakobiformis* (NC_002553), *Marchantia polymorpha* (NC_001660), *Nephroselmis olivacea* (NC_000927), *R. americana* (NC_001823), *Magnetospirillum magnetotacticum* (NZ_AAAP000000000), *Novosphingobium aromaticivorans* (NZ_AAASV000000000), *Rickettsia conorii* (NC_003103). Les séquences de *Rhizopus stolonifer*, de *Malawimonas californiana*, de *H. aroides*, de *J. bahamensis*, de *J. libera*, et de *S. ecuadoriensis* ne sont pas disponibles au public.



CHAPITRE 2 : ARTICLE

Mitochondrial 3' tRNA editing in the jakobid *Seculamonas ecuadoriensis* suggests a eubacterial mechanism for tRNA processing

Jessica Leigh[†] and B. Franz Lang^{*,†}

^{*} Program in Evolutionary Biology, Canadian Institute for Advanced Research

[†] Département de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal (Québec), H3T 1J4, Canada.

Keywords: RNA editing; Jakobids; tRNA processing; tRNA repair; Evolution.

Classification: Biological sciences/Evolution

Corresponding author:

B. Franz Lang

Tel. (514) 343-7936; Fax (514) 343-2210;

E-mail: Franz.Lang@Umontreal.ca

Manuscript information:

Text Pages: 13

Figures: 3

Tables: 2

Word and character counts:

Abstract: 226 words

Manuscript: 33,013 characters

Abbreviations footnote:

Data deposition footnote: *S. ecuadoriensis* sequence data; currently in preparation, will be available at the time of manuscript revision

2.1 Abstract

The jakobid flagellates are bacterivorous protists with mitochondrial genomes that are the most ancestral identified to date, in that they most resemble the genomes of the α -proteobacterial ancestors of the mitochondrion. Because of the bacterial character of jakobid mitochondrial genomes, it was expected that mechanisms for gene expression and RNA structures would be bacterial in nature. However, sequencing of the mitochondrial genome of the jakobid *Seculamonas ecuadoriensis* revealed several apparent mismatches in the acceptor stems of two predicted tRNAs. To investigate this observation, we determined the cDNA sequences of these tRNAs by RT-PCR. Our results show that the last three positions of the 3' extremity plus the discriminator position of seryl and glutamyl tRNAs are altered post-transcriptionally, restoring orthodox base pairing and replacing the discriminator with an adenosine residue, in an editing process that resembles that of the metazoan *Lithobius forficatus*. However, the most 5' of the edited nucleotides is occasionally left unedited, indicating that the editing mechanism proceeds initially by exonucleolytic degradation, followed by repair of the degraded region. This 3' tRNA editing mechanism is likely distinct from that of *L. forficatus*, despite the apparent similarities between the two systems. Furthermore, these results strongly suggest the presence of a tRNA 3' processing mechanism in jakobid mitochondria that resembles that observed in *Escherichia coli*, rather than the processing mechanisms described in the organelles of other eukaryotes.

2.2 Introduction

The jakobid flagellates are bacterivorous, free-living protists, characterized by a ventral groove used for suspension feeding (38). Although the term “jakobids” is often used to refer to both the Malawimonadidae and the Jakobidae, the two groups are phylogenetically distinct, and this term will therefore be used exclusively to refer to the latter family in the following. From an evolutionary perspective, one of the most interesting characteristics of the jakobids is the primitive nature of their mitochondrial genomes (mtDNAs). *Reclinomonas americana*, the best-studied member of the jakobids, has an mtDNA that resembles that of its α -proteobacterial ancestor to a greater extent than any other mtDNA published to date (42). The set of genes encoded by the mitochondrial genome of *R. americana* includes all mitochondrial genes identified in other organisms, as well as 19 additional protein-coding genes. Among these additional genes are several components involved in gene expression, such as the four subunits of a bacterial RNA polymerase (mitochondria in other species use a nucleus-encoded, phage-like RNA polymerase), and the elongation factor Tu. Indeed, many features of the mtDNA of *R. americana*, including a gene order that indicates the presence of bacterial operon-like gene clusters, an RNase P RNA subunit that meets the minimal bacterial consensus, the presence of Shine-Dalgarno motifs (42), and the presence of a derived tmRNA gene (41), indicate that mitochondrial gene expression is bacterial in nature. Complete sequences have recently been determined for the mtDNAs of several other jakobids, including *Seculamonas ecuadoriensis*.

It was expected that the mtDNAs of these species would, similar to *R. americana*, closely follow bacterial mechanisms of gene expression, and that their mitochondrial rRNA and tRNA structures would be similar to those found in bacteria. However, sequencing results from *S. ecuadoriensis* indicate mismatches in the aminoacyl acceptor stems of two tRNA genes, encoding tRNA^{Ser} and tRNA^{Glu}, precluding orthodox tRNA folding. Because proper folding of tRNA molecules is essential for protein synthesis, it was suggested that the transcription products of

these genes must be edited in order to restore proper base pairing of the mature tRNAs.

RNA editing is defined as any modification of an RNA that results in a change in its primary nucleotide sequence relative to that encoded by the corresponding gene (45). This set of processes has been identified in a wide variety of organisms (see, for example, (21) for a recent review). Although RNA editing has predominantly been identified in organelles, it has never been described in prokaryotes. However, cases of RNA editing have also been reported in other systems, such as the mammalian nuclear apolipoprotein B mRNA (46), and the human hepatitis delta virus (47). Editing of tRNAs was initially discovered in the mitochondrion of the amoeboid protist *Acanthamoeba castellanii* (27), where 12 tRNA genes were found to have mismatches in the first three bases of the acceptor stem. The repair of the mismatches by sequence alteration on the 5' half of the acceptor stem has been demonstrated. A similar 5' tRNA editing mechanism was found in chytridiomycete fungi (31).

Some metazoan animals use a different approach to repair tRNA acceptor stem mismatches. In the mitochondria of these animals, a 3' tRNA editing mechanism has been identified. This editing system clearly bears no mechanistic similarity to the systems identified in *A. castellanii* and in the chytrids. In *Gallus gallus*, where mitochondrial tRNA^{Cys} and tRNA^{Tyr} overlap by a single G residue, tRNA^{Tyr} is subject to an editing reaction that yields an A residue in the discriminator position in the place of this G (33). In the squid *Loligo bleekeri* (32) and the land snail *Euhadra herklotsi* (34), mismatched bases in the 3' half of the acceptor stem as well as the discriminator are replaced with A residues to restore orthodox base pairing. These results strongly suggest that 3' tRNA editing in these species involves the action of poly(A) polymerase. In the platypus *Ornithorhynchus anatinus*, a combination of C and A residues are used to replace mismatched bases (48), a result that implicates a modified form of tRNA nucleotidyl transferase (CCase), the enzyme responsible for the addition of CCA 3' to the discriminator position. In *Escherichia coli*, where the 3' terminal CCA is genomically encoded, CCase can repair the CCA

if it is degraded, although it is unable to repair the discriminator or upstream nucleotides (49). The implication of CCase in metazoan mitochondrial 3' tRNA editing is further supported by the results of Reichert and co-workers (18, 36), who demonstrated that overlapping tRNA transcripts in human mitochondria are repaired by human mitochondrial S100 extract *in vitro*. This repair process restores the overlapping discriminator A residue *in vivo*; however, *in vitro*, the extract had a greater affinity for C residues than for A when restoring the discriminator, much as CCase has a greater affinity for C than for A. These authors also demonstrated that the human mitochondrial S100 extract is not only capable of repairing the discriminator position of overlapping tRNAs, but can also repair 3' degraded tRNAs. Although this repair mechanism is not technically a case of RNA editing, it is likely that the two processes are enzymatically related.

In the mitochondria of the centipede *Lithobius forficatus*, a more extensive form of 3' tRNA editing has been identified. In this case, all four bases are used to repair acceptor stem mismatches. Equally interesting in *L. forficatus* is the observation that all 22 tRNAs encoded in the mtDNA would require editing to restore orthodox base-pairing (in 8 of these tRNAs, editing has been demonstrated experimentally). The authors suggest that 3' tRNA editing in *L. forficatus* mitochondria is mechanistically distinct from the editing mechanisms described in mitochondria of other animals, in that it must use an RNA-dependent RNA polymerase (RdRp), instead of a template-independent CCase-like enzyme (3).

In this article, we describe a 3' tRNA editing mechanism in the mitochondria of *S. ecuadoriensis* that superficially resembles that identified in *L. forficatus*, though the two systems have certainly evolved independently. Because RNA editing has generally been identified in the more derived organelles of eukaryotes, the discovery of such a mechanism in the mitochondria of a jakobid was completely unexpected. However, we will demonstrate that the particular details of the 3' tRNA editing mechanism in *S. ecuadoriensis* mitochondria suggest that it has evolved as an adaptation of a eubacterial 3' tRNA processing system, reminiscent of that observed in *E. coli*.

2.3 Materials and Methods

2.3.1 Cell culture, cloning, sequencing, and annotation of mtDNA

DNA sequences of *S. ecuadoriensis* were determined under the auspices of the Organelle Genome Megasequencing Project. Details of growth conditions and mtDNA isolation and cloning will be presented elsewhere in conjunction with a more detailed description of the complete sequence (Lang *et al.* unpublished results). DNA sequencing, data entry, and sequence analysis were performed as described (50). tRNA genes were identified using tRNAscan-SE (51).

2.3.2 Phylogenetic analysis

Amino acid sequences inferred from mitochondrial genes encoding the proteins of the cytochrome C oxidase complex (Cox1, Cox2, and Cox3) and apocytochrome b (Cob) were concatenated and aligned using ClustalW (52). Ambiguous portions of the alignment were removed using SOAP (53), resulting in a total of 1197 retained amino acid positions. A maximum likelihood (ML) phylogeny was inferred using ProML. Six categories were used under a discrete gamma distribution model, with an alpha value of 0.64 (calculated using CodeML; (44)). ML bootstraps were calculated using Seqboot to generate bootstrapped data sets, ProML (as described above) to infer trees from these sets, and finally Consense to determine the extended majority rule consensus tree. ProML, Seqboot, and Consense are provided with the PHYLIP package (Felsenstein, J., 2002; v. 3.6a 2.1, distributed by the author).

2.3.3 RNA purification

Cells were cultured in WCL medium, buffered with 20 mM HEPES, harvested and disrupted with glass beads, and enriched for mitochondria by differential centrifugation. Mitochondria were lysed with SDS and proteinase K and enriched for mitochondrial tRNAs by centrifugation through a glycerol cushion. The

upper tRNA-containing fractions were pooled, and protein was eliminated by repeated phenol-chloroform extractions, followed by multiple ethanol precipitations.

2.3.4 Oligonucleotides

The following primer sequences were used for RT-PCR:

tRNA amplified	Direction	Sequence
Alanine	Forward	CATATGCTCTGCCAAATGAGCT
Alanine	Reverse	TTTTGCACGCATAGGGTAACCA
Cysteine	Forward	CCCATACATTGCCATTATGTTA
Cysteine	Reverse	CTGCAAATCCTATAATGGCG
Glutamate	Forward	CGATGTCCTAACCCTAGACGA
Glutamate	Reverse	TCATGGCGTAAACACGGGTTC
Histidine	Forward	CACTCTAACCAACTGAGTTA
Histidine	Reverse	TTGTGATTCTGGAAGTCACG
Serine	Forward	GCGCAATAGACCACTCTGCCA
Serine	Reverse	TTTGGAATTCGTGTTAATCTAAT

2.3.5 RT-PCR, cloning, and sequencing

tRNAs were circularized and cDNAs synthesized following the protocol of Yokobori and Pääbo (34), with the following adaptations: the tRNA concentration used for circularization was reduced to 40 µg/mL, and the circularized tRNA concentration used for reverse transcription was reduced to 25 µg/mL. The resulting cDNAs were amplified by PCR. PCR products were incubated with a mixture of T7 DNA polymerase and *Escherichia coli* DNA polymerase I (Klenow fragment) to generate blunt ends, and the PCR fragment of the appropriate size was recovered after agarose gel electrophoresis. The purified PCR fragment was cloned into the *Eco*R5 cloning site of the phagemid pFBS (B.F. Lang, unpublished) and sequenced.

2.4 Results and Discussion

2.4.1 Phylogenetic analysis indicates that *S. ecuadoriensis* and *R. americana* are monophyletic

A global eukaryotic phylogeny, shown in Figure II.1, was constructed from mitochondrion-encoded protein sequences using ML in order to confirm the placement of *S. ecuadoriensis* within the jakobids. ML bootstrap values were calculated from 100 replicates.

Because only 1197 amino acid positions were used in this analysis, many features of this topology are poorly supported. However, there is strong support for the monophyly of the jakobids (100%), as well as of the Opisthokonts (fungi plus animals, 99%). Clearly, the phylogenetic relationship between *Seculamonas* and the metazoans is very distant. Other groups that are well-supported even with so few data are the plants plus green algae and the red algae (both 100%).

2.4.2 Editing of mitochondrial seryl and glutamyl tRNAs in *S. ecuadoriensis*

We have identified 28 tRNA genes in the mtDNA of *S. ecuadoriensis*. Of these, the genes for tRNA^{Ser}(GGA) and tRNA^{Glu}(UUC) encode 2 and 3 mismatches, respectively, in the acceptor stem region. In addition, tRNA^{Cys}(GCA) encodes a terminal G:U pair in the acceptor stem and tRNA^{Ala}(UGC) contains two G:U pairs in the acceptor stem region, one of which (between G₃ and U₇₀) is a highly conserved structural determinant of this tRNA (54), although this position is edited in the 5' portion of the acceptor stem to produce an A:U pair in the mitochondria of *A. castellanii* (28). Among the remaining tRNA genes, several acceptor stems contain G:U pairs, and some contain U:U pairs at positions near the base of the helix. These findings are summarized in Table 1.

Figure II.1: Global eukaryotic phylogeny

This tree was inferred from mitochondrion-encoded protein sequences of the genes *cob*, *cox1*, *cox2*, and *cox3* (a total of 1197 amino acid positions) by maximum likelihood with ProML (Felsenstein 2002). ML bootstraps were calculated from 100 replicates. Only bootstrap values greater than 50% are indicated. Sequences were obtained from Genbank: *Monosiga brevicollis* (NC_004309), *Metridium senile* (NC_000933), *Allomyces macrogynus* (NC_001715), *Dictyostelium discoideum* (NC_000895), *Acanthamoeba castellanii* (NC_001637), *Rhodomonas salina* (NC_002572), *Phytophthora infestans* (NC_002387), *Chrysodidymus synuroideus* (NC_002174), *Nephroselmis olivacea* (AF110138), *Marchantia polymorpha* (NC_001660), *Prototheca wickerhamii* (NC_001613), *Porphyra purpurea* (NC_002007), *Chondrus crispus* (NC_001677), *Seculamonas ecuadoriensis* (****), *Reclinomonas americana* (NC_001823), *Malawimonas jakobiformis* (NC_002553), *Rickettsia conorii* (NC_003103), *Novosphingobium aromaticivorans* (NZ_AAAV000000000), *Magnetospirillum magnetotacticum* (NZ_AAAP000000000). Sequences from *Rhizopus stolonifer* are available on <http://megasun.bch.umontreal.ca/People/lang/FMGP/proteins.html>.

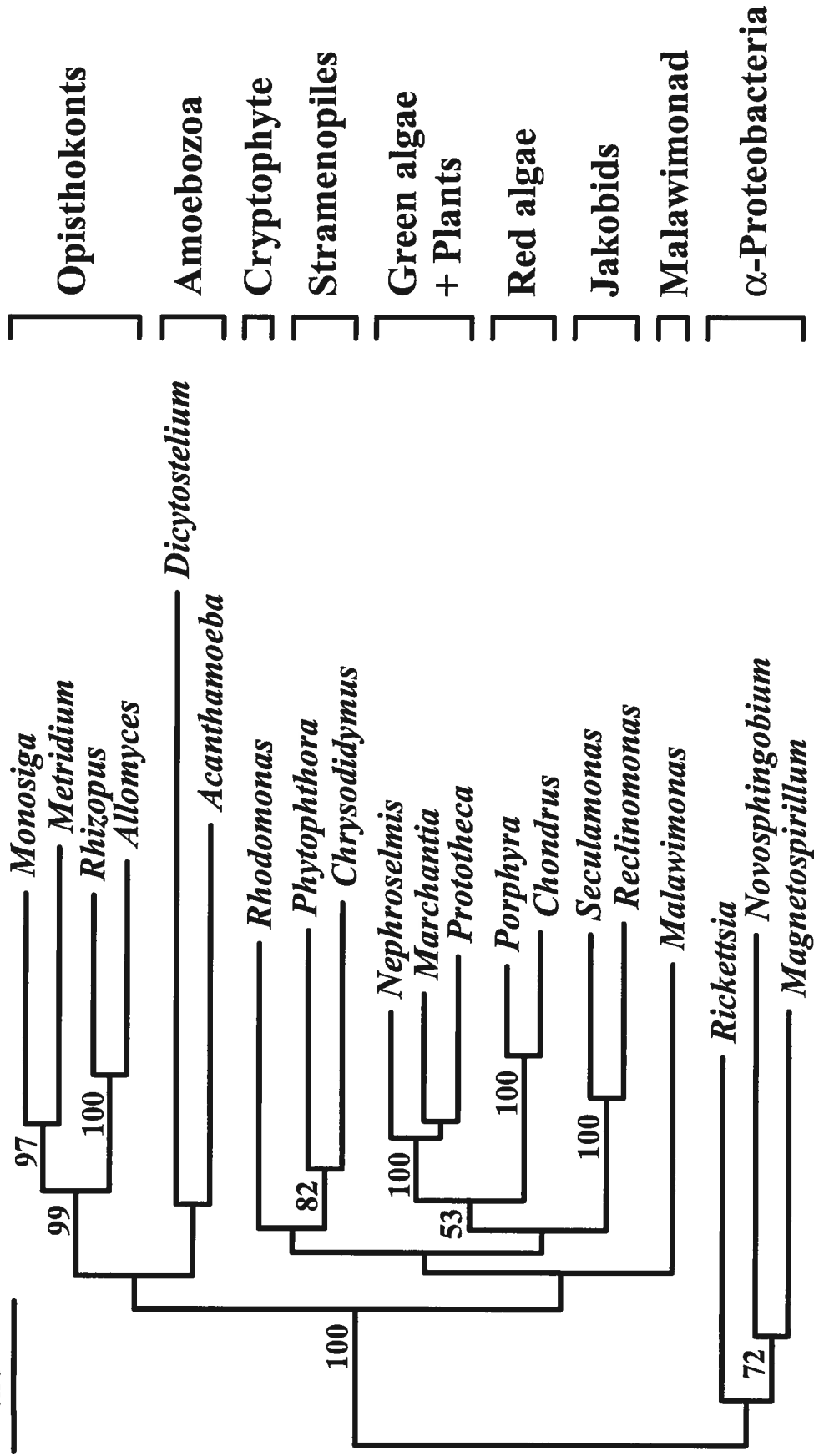


Table 1: Non-Watson-Crick base pairs in aminoacyl acceptor stems

Shown here is a summary of unorthodox base pairs in the acceptor stems of inferred tRNAs in the mtDNA of *S. ecuadoriensis*. tRNAs are identified by one-letter amino acid codes, followed by the anticodon sequence. Positions of mispairings are given as 5' base position: 3' base position, and are numbered according to standard nomenclature. With the exception of those found in tRNA^{Ser} and tRNA^{Glu}, all unusual base pairs are G:U or U:U pairs. Acceptor stem mismatches in these two tRNAs are not indicated in this table, but are shown in Figure II.2.

Table 1: Non-Watson-Crick base pairs in aminoacyl acceptor stems

G:U pairs		U:U pairs	
tRNA identity	Mismatch position	tRNA identity	Mismatch position
A(UGC)	3:70, 5:68	H(GUG)	5:68
C(GCA)	1:72	L(GAG)	6:67
F(GAA)	7:66	M(CAU) ^e	6:67
G(GCC)	7:66	P(UGG)	6:67
K(UUA)	2:71		
M(CAU) ^e	2:71		
N(GUU)	3:70		
V(UAC)	2:71		

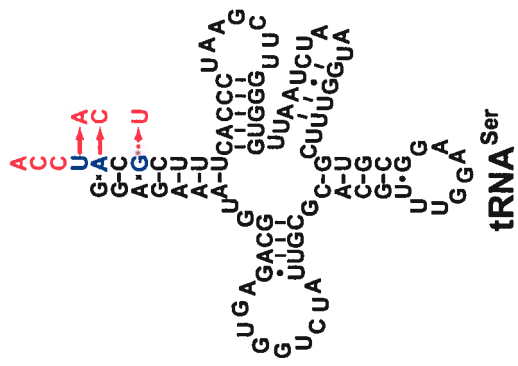
Although G:U pairs are not uncommon in tRNA stems, the presence of other, less stable mispairings suggested that they might be repaired by an editing mechanism. In order to determine whether the mismatches in seryl, glutamyl, cysteinyl, and alanyl tRNA acceptor stems are edited, cDNA sequences of these molecules were determined using RT-PCR. For these tRNAs, 25, 27, 11, and 20 independent clones were sequenced, respectively. In addition, the sequence of tRNA^{His} was determined in order to verify whether its discriminator is incorporated in the 8 base pair acceptor stem that is characteristic of tRNA^{His} structures (in *L. forficatus*, following editing, tRNA^{His} has an additional non-canonical 3' discriminator in addition to the 8 base pair acceptor stem; (3). In *S. ecuadoriensis*, only three tRNA^{His} molecules were sequenced.

Our results indicate that the sequences of tRNA^{Ser} and tRNA^{Glu} differ from those of their corresponding genes. Inferred secondary structures of these tRNAs, indicating sites of editing, are shown in Figure II.2. The mismatched nucleotides located in the 3' portion of the acceptor stem of these tRNAs were found to be replaced with nucleotides that form orthodox, Watson-Crick base pairs with the corresponding nucleotides in the 5' portion of the acceptor stem. In both cases, the altered residues were restricted to the three final residues of the 3' end of the acceptor stem. In addition, the discriminator position was replaced by an A in both cases. Because CCA was found to be added to both seryl and glutamyl tRNAs, they are presumably aminoacylated *in vivo*. These results demonstrate that a 3' tRNA editing system is present in *S. ecuadoriensis* mitochondria.

Also shown in Figure II.2 are inferred secondary structures of tRNA^{Ala}, tRNA^{His}, and tRNA^{Cys}, for which the RNA-level sequences were found to be identical to the genomic sequence. Additionally, tRNA^{His} was found to have the usual 8 base pair acceptor stem, into which the discriminator was incorporated. Although these three tRNAs all contain mispairings in the acceptor stem region, the destabilizing effect of G:U pairs (in the case of tRNA^{Ala} and tRNA^{Cys}) and of relatively deep U:U pairs (in the case of tRNA^{His}) appears to be too weak to require editing. All three of these tRNA molecules were found to terminate with the 3' CCA.

Figure II.2: Inferred tRNA secondary structures

tRNA secondary structures were inferred from primary, mtDNA sequences. Shown here are secondary structures for tRNA^{Ser}, tRNA^{Glu}, tRNA^{Ala}, tRNA^{His}, and tRNA^{Cys}. The genomic sequence is shown in black, with mismatches and unusual discriminators shown in blue. Nucleotide identities following editing are indicated in red. Red arrows indicate editing; at partially edited positions, arrows have dotted lines.



2.4.3 3' tRNA editing suggests a eubacterial mechanism for tRNA processing

In all sequenced clones, tRNA^{Ser} and tRNA^{Glu} were found to be edited. However, in some cases for both tRNA^{Ser} and tRNA^{Glu}, the furthest upstream of the mismatched nucleotides (G₇₀ and A₇₀, respectively, according to standard nomenclature) was found to have the mtDNA-encoded identity (G or A), rather than the edited identity (U or G); in other words, this position is only partially edited. The partially edited positions are indicated in Figure II.2. Table 2 summarizes the totals for unmodified versus modified first position identity. The percentage of sequenced tRNA^{Ser} molecules that was unmodified at this position was found to be 26%, while the percentage of partially edited tRNA^{Glu} molecules was found to be 28%. These percentages were calculated based on the total of completely and partially edited tRNA molecules.

The presence of these partially edited tRNA molecules strongly suggests that the mechanism for 3' tRNA editing in *S. ecuadoriensis* mitochondria involves exonucleolytic degradation of the 3' half of the mismatched region of the acceptor stem, followed by its resynthesis (Figure II.3a).

This proposed mechanism is reminiscent of tRNA 3' processing in eubacteria (Figure II.3b). Studies of 3' tRNA processing have primarily been carried out in *Escherichia coli*, in which processing begins with an endonucleolytic cleavage of the primary tRNA transcript downstream of the region of the tRNA. The transcript is then trimmed by exonucleases, most notably RNase PH and RNase T. During the course of this trimming process, the 5' extremity is cleaved by RNase P. In the case of *E. coli*, the 3' terminal CCA is encoded in all tRNA genes, and the enzyme CCCase serves the purpose of repairing the CCA if the exonucleolytic trimming reaction proceeds into the tRNA region of the transcript. Thus, tRNA 3' processing is a competition between exonucleases and CCCase, a contest that ends when the tRNA is protected from degradation with the addition of an amino acid (6).

3' tRNA editing in the mitochondria of *S. ecuadoriensis* likely proceeds *via* an adaptation of such a mechanism. We propose that mitochondrial 3' tRNA

Figure II.3: Processing, editing, and repair of tRNAs

a) Mitochondrial 3' tRNA editing in *S. ecuadoriensis*. Exonucleolytic degradation (blue) of the 3' trailer and any mismatched nucleotides in the 3' region of the acceptor stem. During this exonucleolytic process, RNase P (red) processes the 5' region of the molecule. The degraded 3' region of the acceptor stem is repaired by a CCCase-like enzyme (green). Finally, the tRNA is aminoacylated, protecting it from further degradation, and allowing it to function in translation. b) Processing of tRNAs in *E. coli*. Exonucleases (blue) begin trimming the 3' trailer region. When the 3' trailer is sufficiently short, RNase P (red) cleaves off the 5' leader sequence. Exonucleolytic degradation of the 3' trailer continues, yielding a 3' region of the acceptor stem, complete with genome-encoded CCA (6, 12). c) Editing and repair in metazoan mitochondria (dynamic model for tRNA repair). Mismatched acceptor stems are susceptible to 3' exonucleolytic degradation, which is followed by random incorporation of nucleotides, likely catalyzed by a CCCase-like enzyme. If the nucleotides incorporated are able to form orthodox base-pairs with the 5' portion of the acceptor stem, the structure is aminoacylated, and protected from further degradation. If the nucleotides incorporated cannot form Watson-Crick base-pairs, no aminoacylation occurs, and the 3' region is degraded again, then re-synthesized. This dynamic process continues until orthodox base pairs form, and the structure is protected by aminoacylation (18).

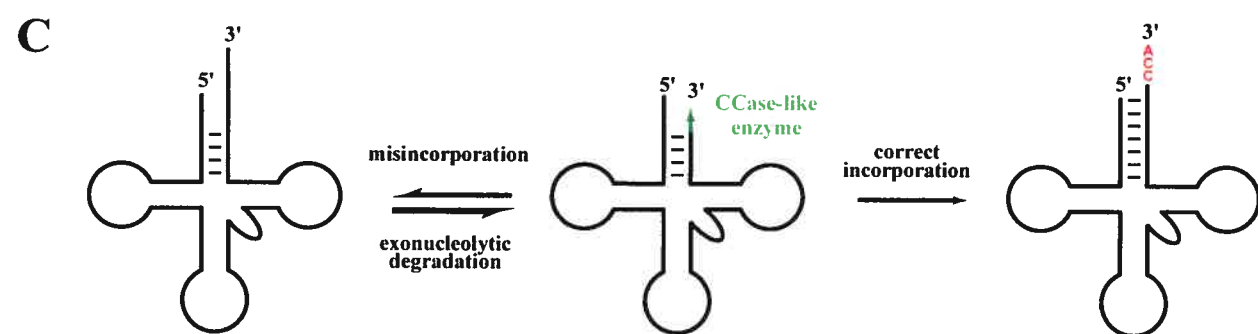
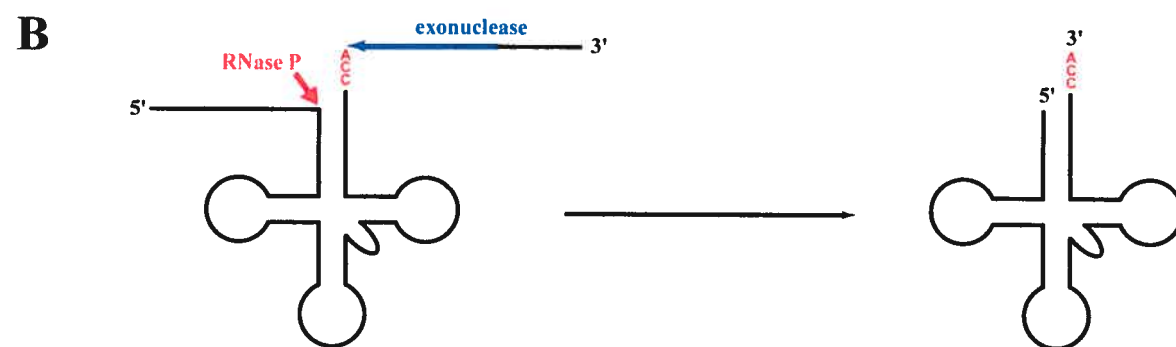
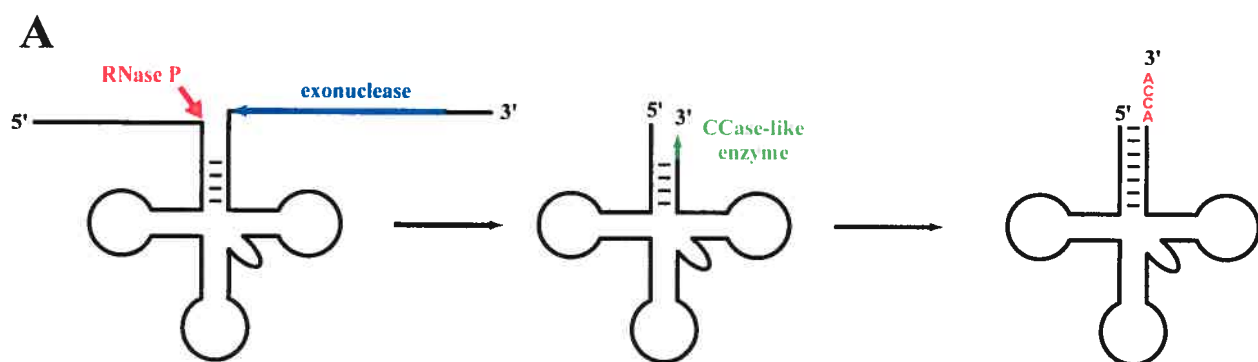


Table II.2: Summary of tRNA sequencing results

Shown here are total numbers of tRNA molecules sequenced. Also indicated are numbers of fully edited, partially edited, and unedited molecules for each tRNA species studied. Editing was only found in tRNA^{Ser} and tRNA^{Glu}, and in both cases, a number of partially edited molecules (i.e., unedited at the most upstream nucleotide) was found. tRNA^{Ala}, tRNA^{His}, and tRNA^{Cys} were found to match the genomic sequence, although CCA was added to the molecule.

Table 2: Summary of sequencing results

tRNA	Total # sequenced	Completely edited	Partially edited	Unedited	Other	% Partially edited	% Unedited
Alanine	20	0	0	19	1	0	100%
Cysteine	11	0	0	11	0	0	100%
Glutamate	27	13	5	0	9	28%	0%
Serine	25	17	6	0	2	26%	0%
Histidine	3	0	0	3	0	0	100%

processing in *S. ecuadoriensis* involves an endonucleolytic cleavage, followed by exonucleolytic trimming. This nucleolytic activity is in competition with a repair enzyme, possibly a modified CCase, that regenerates the degraded 3' end of the tRNA. This model is in contrast to the tRNA processing pathways of most organelles, which are thought to resemble nuclear tRNA processing pathways to a greater extent than those of their eubacterial ancestors. For example, in *Saccharomyces cerevisiae*, 3' end maturation results from a single endonucleolytic cleavage (16); in the nucleus, the 3' trailer is also removed by an endonuclease *in vivo* (10). Similarly, in plants, mature 3' ends of nuclear, mitochondrial, and chloroplast tRNAs are generated by a single endonucleolytic cleavage, catalyzed by RNase Z (13, 14, 55).

In metazoans, mitochondrial size constraints result in tRNA genes that are generally separated from downstream RNA or protein-coding genes by very few nucleotides; in some cases, tRNA genes even overlap. Consequently, most tRNAs seem to be processed by a single endonucleolytic cleavage that recognizes the upstream tRNA (17, 36). However, it is likely that an exonuclease that can act on the 3' end of tRNAs is present, as animal mitochondria have a repair activity that is able to regenerate the 3' extremity of tRNAs. The dynamic model of tRNA repair (Figure II.3c; 18), which explains the nature of this process, suggests that this repair process is template-independent, beginning with random nucleotide insertion by a CCase-related enzyme. If the incorrect nucleotides are incorporated, the acceptor stem is unable to form orthodox base pairs, is not recognized by the aminoacyl tRNA synthetase, and is susceptible to exonucleolytic degradation. Once the correct nucleotides have been inserted, the acceptor stem base pairing is permitted, and the tRNA is aminoacylated, protecting it from further degradation. Schürer *et al.* (12) suggested that such a model might explain 3' tRNA repair and editing in all metazoan mitochondria, including *L. forficatus*. However, we feel that the original hypothesis of Lavrov *et al.* (3), by which the resynthesis of the 3' region of the acceptor stem is template-directed, is more likely, given that all four nucleotides are used to edit five positions (including the discriminator) in all 22 tRNAs, which would make random nucleotide incorporation extremely inefficient, if not impossible.

Template-independent resynthesis of the degraded region of the acceptor stem would seem similarly inefficient in *S. ecuadoriensis* due to the 256 possible sequences of the tetranucleotide formed by the edited positions. However, the insertion of nucleotides is not necessarily truly random if a CCase-related enzyme is responsible for resynthesis, due to the preference of CCase for C and A residues (56). In the two edited tRNAs in *S. ecuadoriensis*, all inserted nucleotides at the six completely edited positions are either C or A (although one G and one U are inserted, these are in the partially edited positions).

If the dynamic model for tRNA repair can explain 3' tRNA editing in *S. ecuadoriensis*, we would expect to find some intermediate tRNA molecules, in which the 3' end of the molecule was incorrectly repaired. Although none of these intermediates were observed, there were several cases in which the 3' end of the molecule was badly degraded, while the 5' end was intact (results not shown).

Another possibility is that 3' tRNA editing in *S. ecuadoriensis* is truly template-dependent, as Lavrov *et al.* (3) suggested was the case for *L. forficatus*. Such would preclude the involvement of CCase, which is a strictly template-independent enzyme. Template-directed editing of the 3' end of the acceptor stem would require a 5'-to-3' RdRp. Such a protein is encoded by the genomes of many RNA viruses, as well as by the mitochondrial genome of *Arabidopsis thaliana* and by an RNA element found in the mitochondria of the fungus *Ophiostoma novo-ulmi* (57). An RdRp, possibly laterally transferred from a virus, could be responsible for editing in *S. ecuadoriensis*. In this case, the discriminator position might be added in a template-independent fashion by poly(A)polymerase. However, we feel that the suggestion that an adapted CCase carries out the template-independent resynthesis of the 3' portion of the acceptor stem is more likely, given the strong bias towards the insertion of C and A residues during editing.

A final possibility is that 3' tRNA editing in *S. ecuadoriensis* involves base modification, rather than deletion/insertion. However, during editing, various positions are converted from purines to pyrimidines, as well as from pyrimidines to

purines. Clearly, amination and deamination reactions would be insufficient to account for these changes, and other types of modification would be much more complex, and therefore unlikely.

2.4.4 *S. ecuadoriensis* mitochondrial 3' tRNA editing analogous to *L. forficatus*?

Editing mechanisms affecting the 3' region of the acceptor stem of mitochondrial tRNAs have previously been identified in animals. Although the mitochondrial 3' tRNA editing mechanisms of many metazoans use only A, or A and C residues, to repair acceptor stem mismatches (32-34, 48), the editing system of *S. ecuadoriensis* particularly resembles that of *L. forficatus*, in which all four nucleotides are used to replace mismatched bases of the 3' end of the acceptor stem, including the discriminator position (3).

However, despite the apparent similarity between the 3' tRNA editing systems of *S. ecuadoriensis* and *L. forficatus*, the phylogenetic relationship between these two species is so distant (Figure II.1) that this similarity is clearly an example of convergent evolution. If 3' tRNA editing were present in the common ancestor of the metazoa and the jakobids, we would expect to find evidence in the form of apparent mismatches in mitochondrial tRNA acceptor stems in other eukaryotes, such as all metazoans, fungi, and jakobids. However, there is no evidence that 3' tRNA editing was present even in the common ancestor of the jakobids; in *R. americana*, a close examination of predicted tRNAs reveals no significant mismatches.

In addition, given that only two tRNAs appear to be modified post-transcriptionally in *S. ecuadoriensis*, it is likely that this editing system evolved relatively recently, as the presence of such a system would virtually eliminate selection pressure on the edited region of the molecule, allowing the sequence to diverge freely. In contrast, the editing system in *L. forficatus* is likely a much older adaptation, since editing has been demonstrated experimentally or inferred, based on sequence data for all 22 tRNAs encoded in the mtDNA (3).

Finally, the two editing systems appear to be mechanistically distinct. The preferential use of A and C residues seen in 3' editing of *S. ecuadoriensis* mitochondrial tRNAs is far less significant in *L. forficatus*, where C and A make up only 62% of nucleotides used in editing of the 8 tRNAs in which editing was demonstrated experimentally (including the discriminator, which was always replaced with an A). Additionally, no partial editing of tRNAs was found in *L. forficatus* (3). If *S. ecuadoriensis* and *L. forficatus* shared an analogous exonucleolysis-resynthesis mechanism for 3' tRNA editing, it is likely that at least some cases of partial editing would have been observed. Perhaps the mismatched region of the acceptor stem is removed by an endonucleolytic process, resulting in a 3' truncated tRNA that is then repaired. This model is supported by the results of Rossmanith *et al.* (17), who demonstrated that, in human mitochondria, maturation of the 3' region of tRNAs is the result of an endonucleolytic cleavage.

2.5 Acknowledgements

We wish to thank T. Nerad (ATCC), who isolated *S. ecuadoriensis*. We also thank L. Forget for library construction and DNA sequencing, Z. Wang and L. Forget for DNA sequencing, and D. Lavrov and C. Bullerwell for critical reading of the manuscript. This work was supported by the 'Canadian Institutes of Health Research' (CIHR, MSP-14226). B.F.L. is Imasco Fellow in the Program of Evolutionary Biology of the Canadian Institute for Advanced Research (CIAR), whom we thank for salary and interaction support.

2.5 References

Les références ont été intégrées avec celles qui se trouvent dans la bibliographie à la fin de ce mémoire.

CHAPITRE 3 : DISCUSSION ET CONCLUSION

3.1 Discussion

3.1.1 Édition des ARNt chez *Seculamonas ecuadoriensis*

L'édition des ARN, un ensemble de divers processus qui changent la séquence primaire d'un ARN comparativement à celle codée par le gène correspondant, a été mise en évidence dans une grande variété d'espèces. Chez les plantes, ces événements sont généralement observés dans les organites, surtout dans la mitochondrie, où jusqu'à 1000 sites de désaminations peuvent changer le message codé par un ARNm ou la conformation d'un ARN non-traduit. Chez les virus et dans le noyau des animaux, l'édition produit souvent deux ou plusieurs protéines à partir d'un seul gène. Dans la mitochondrie des animaux, l'édition effectue la réparation des ARNt.

Il est surprenant que *S. ecuadoriensis*, un jakobide, ait un système d'édition des ARNt mitochondriaux. L'édition des ARN n'a jamais été mise en évidence chez les bactéries, et l'ADNmt des jakobides est plus bactérien de caractère que toute autre mitochondrie étudiée jusqu'à présent. Dans ces génomes, la présence de plusieurs composantes des systèmes d'expression des gènes qui n'ont jamais été observés dans les mitochondries en dehors des jakobides indique que l'expression des gènes mitochondriaux dans cette lignée ressemble probablement beaucoup à celle des bactéries.

Cependant, nos résultats démontrent clairement qu'il existe un mécanisme d'édition des ARNt mitochondriaux chez *S. ecuadoriensis*. Deux des ARNt codés par l'ADNmt de cet organisme sont édités dans la région de l'extrémité 3' de la tige acceptrice. Dans les deux cas, les positions éditées sont limitées aux quatre dernières positions de cette extrémité, y compris le discriminateur. Le CCA terminal est rajouté aux deux molécules, ce qui indique qu'elles sont fonctionnelles.

3.1.2 Évolution récente du mécanisme d'édition

Une des considérations intéressantes lors de la découverte d'un mécanisme exceptionnel, comme un nouveau mécanisme d'édition des ARN, est l'estimation du temps où ce mécanisme a évolué. Dans plusieurs cas de l'édition des ARN, une date d'origine du mécanisme peut être déduite en se basant sur sa distribution phylogénétique. Par exemple, Horton et Landweber (22) ont fait des analyses phylogénétiques sur les myxomycètes pour pouvoir estimer où, dans la phylogénie de ce groupe, les différentes formes d'édition des ARN ont été développées.

Dans le cas de *S. ecuadoriensis*, l'apparition de l'édition des ARNt mitochondriaux suit évidemment la divergence entre cet organisme et les autres jakobides pour lesquels l'ADNmt a été séquencé. Chez *Jakoba libera*, *J. bahamensis*, et *R. americana*, les séquences génomiques des ARNt mitochondriaux ne suggèrent pas l'édition : sauf la présence de quelques paires de bases GU, les régions des extrémités forment des paires de bases de type Watson-Crick.

La présence de seulement deux ARNt édités dans la mitochondrie de *S. ecuadoriensis* indique aussi une origine récente pour ce mécanisme d'édition. Un tel système réduit la pression de sélection qui maintient les paires de bases entre les extrémités des ARNt. Si les quatre derniers nucléotides de l'extrémité 3' peuvent être facilement réparés, il n'est plus nécessaire que leur identité soit conservée. Conséquemment, il est probable que la séquence de l'extrémité 3' de tous les ARNt mitochondriaux soit éventuellement érodée, comme il a été observé chez *L. forficatus*, ou dans la séquence de l'extrémité 5' de *A. castellanii*.

3.1.3 Modèle théorique pour le mécanisme d'édition

L'identification d'un mécanisme d'édition exige la discussion des mécanismes d'action possible. L'édition des ARN peut être causé par de nombreuses activités enzymatiques : des délétions, des insertions, et des modifications ont toutes été impliquées dans l'édition des ARN.

La fréquence assez élevée d'édition partielle à la troisième position de la tige acceptrice indique que l'édition des ARNt mitochondriaux chez *S. ecuadoriensis* commence probablement par une dégradation exonucléolytique qui n'est pas toujours complétée. Ceci diffère de l'édition de l'extrémité 3' des ARNt mitochondriaux chez la plupart des animaux (3, 32-36), où les ARNt sont nécessairement séparés par un clivage endonucléolytique, et sont ensuite réparés.

La dégradation de la région mésappariée de l'extrémité 3' est sans doute suivie de sa resynthèse. Cependant, la resynthèse pourrait être soit dirigée par l'extrémité 5', soit indépendante. Si l'extrémité 5' sert de matrice pour la synthèse de l'extrémité 3', comme Lavrov et ses collègues (3) ont suggéré dans le cas du centipède *L. forficatus*, une ARN polymérase ARN-dépendante (RnRp) serait nécessaire. Une telle enzyme existe chez plusieurs virus, et son homologue peut être adopté par la mitochondrie suite à un événement de transfert latéral. Il est possible, alors, que *S. ecuadoriensis* ait acquis une RnRp. Mais nous préférons la deuxième option. Si l'extrémité 3' est synthétisée de façon matrice-indépendante, une CCCase modifiée pourrait être l'activité enzymatique responsable.

3.1.4 Implications pour la maturation de l'extrémité 3' des ARNt

La compétition entre la CCCase et la dégradation exonucléolytique que ce modèle implique ressemble beaucoup à la maturation des ARNt chez *E. coli*, où le rôle de la CCCase est de réparer le CCA terminal quand les exonucléases responsables pour la maturation de l'extrémité 3' agissent trop loin.

Nos résultats indiquent que, chez *S. ecuadoriensis*, il y a une compétition entre l'enzyme qui digère l'extrémité 3' et l'enzyme de synthèse. Cependant, il reste à démontrer la nature de l'enzyme de synthèse. Une CCCase nécessiterait deux modifications : il lui faudrait la capacité d'ajouter n'importe lequel des quatre nucléotides; et il lui faudrait une poche de liaison au substrat assez grande pour pouvoir accommoder beaucoup plus de la tige acceptrice que la CCCase standard.

3.1.5 Modèle théorique pour l'évolution de l'édition des ARNt chez *S. ecuadoriensis*

Les différentes formes d'édition des ARN peuvent être divisées en deux catégories fonctionnelles. Dans plusieurs cas, l'édition est utilisée pour réparer un transcrit non-fonctionnel. Cette catégorie inclut l'édition des ARNt observée chez les métazoaires et chez *S. ecuadoriensis*, ainsi que l'édition des ARNm chez les kinétoplastides et l'édition des ARN mitochondriaux et chloroplastiques. Tandis que cette forme d'édition est plutôt correctionnelle, l'édition peut aussi être utilisée pour produire un ou plusieurs produits à partir d'un seul gène. Incluses dans cette catégorie sont l'édition de l'ARNt^{Gly} chez les marsupiaux, l'édition de l'ARNm de la apoB dans le noyau des mammifères, et l'édition du transcrit du gène P chez les paramyxovirus.

L'évolution de l'édition des ARN de cette deuxième catégorie fonctionnelle est facile à comprendre, car elle augmente la capacité codante du génome, ce qui pourrait évidemment être sélectionné. Mais comment expliquer la première catégorie? Dans le cas de l'édition des ARNt chez *S. ecuadoriensis*, le mécanisme d'édition a probablement commencé comme mécanisme de réparation. Comme le CCA terminal n'est pas codé par le génome mitochondrial (comme chez *E. coli*), ce mécanisme serait nécessaire si la maturation de l'extrémité 3' était dirigée par des exonucléases, comme nous le proposons. Le résultat d'un tel mécanisme serait une relaxation des contraintes sur la séquence de cette extrémité et, éventuellement, sa divergence. Le mécanisme d'édition des ARNt mitochondriaux chez *S. ecuadoriensis* ressemble à celui identifié chez le centipède *L. forficatus*, mais la présence de l'édition chez cette espèce est clairement un exemple de l'évolution convergente.

3.2 Conclusion

Les jakobides, un groupe de protistes biflagellés, ont un ADNmt qui ressemble à un mini-génome bactérien. Ils contiennent plusieurs gènes d'origine bactérienne dont les produits servent à l'expression des gènes et qui ne sont pas trouvés ailleurs chez les eucaryotes, et une organisation qui ressemble aux opérons bactériens. *S. ecuadoriensis* n'est pas une exception. Pour cette raison, on s'attendrait à ce que l'expression des gènes mitochondriaux dans cette espèce soit de type bactérien.

Cependant, chez *S. ecuadoriensis*, il existe un mécanisme d'édition des ARNt mitochondriaux, un phénomène qui n'a jamais été observé chez les bactéries, mais qui est plutôt associé aux eucaryotes. Deux des ARNt mitochondriaux, ARNt^{Ser} et ARNt^{Glu}, subissent des éditions dans les trois dernières positions de l'extrémité 3' de la tige acceptrice. Malgré qu'il soit surprenant qu'il existe un tel mécanisme chez *S. ecuadoriensis*, nous croyons que sa présence indique un mécanisme de type bactérien de maturation des ARNt chez les jakobides.

BIBLIOGRAPHIE

1. Brennicke, A., Marchfelder, A. & Binder, S. (1999) *FEMS Microbiol Rev* **23**, 297-316.
2. Beagley, C. T., Okimoto, R. & Wolstenholme, D. R. (1999) *Genetics* **152**, 641-52.
3. Lavrov, D. V., Brown, W. M. & Boore, J. L. (2000) *Proc Natl Acad Sci U S A* **97**, 13738-42.
4. Li, Z. & Deutscher, M. P. (1994) *J Biol Chem* **269**, 6064-71.
5. Kelly, K. O., Reuven, N. B., Li, Z. & Deutscher, M. P. (1992) *J Biol Chem* **267**, 16015-8.
6. Li, Z. & Deutscher, M. P. (1996) *Cell* **86**, 503-12.
7. Reuven, N. B. & Deutscher, M. P. (1993) *FASEB J* **7**, 143-8.
8. Zhu, L. & Deutscher, M. P. (1987) *EMBO J* **6**, 2473-7.
9. Papadimitriou, A. & Gross, H. J. (1996) *Eur J Biochem* **242**, 747-59.
10. Yoo, C. J. & Wolin, S. L. (1997) *Cell* **89**, 393-402.
11. Mörl, M. & Marchfelder, A. (2001) *EMBO Rep* **2**, 17-20.
12. Schürer, H., Schiffer, S., Marchfelder, A. & Mörl, M. (2001) *Biol Chem* **382**, 1147-56.
13. Mayer, M., Schiffer, S. & Marchfelder, A. (2000) *Biochemistry* **39**, 2096-105.
14. Schiffer, S., Rösch, S. & Marchfelder, A. (2002) *EMBO J* **21**, 2769-77.
15. Schierling, K., Rösch, S., Rupprecht, R., Schiffer, S. & Marchfelder, A. (2002) *J Mol Biol* **316**, 895-902.
16. Chen, J. Y. & Martin, N. C. (1988) *J Biol Chem* **263**, 13677-82.
17. Rossmanith, W., Tullo, A., Potuschak, T., Karwan, R. & Sbisà, E. (1995) *J Biol Chem* **270**, 12885-91.
18. Reichert, A. S. & Mörl, M. (2000) *Nucleic Acids Res* **28**, 2043-8.
19. Oommen, A., Li, X. Q. & Gegenheimer, P. (1992) *Mol Cell Biol* **12**, 865-75.
20. Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H. & Tromp, M. C. (1986) *Cell* **46**, 819-26.

21. Gott, J. M. & Emeson, R. B. (2000) *Annu Rev Genet* **34**, 499-531.
22. Horton, T. L. & Landweber, L. F. (2000) *RNA* **6**, 1339-46.
23. Horton, T. L. & Landweber, L. F. (2002) *Curr Opin Microbiol* **5**, 620-6.
24. Maslov, D. A. & Simpson, L. (1992) *Cell* **70**, 459-67.
25. Hausmann, S., Jacques, J. P. & Kolakofsky, D. (1996) *RNA* **2**, 1033-45.
26. Rueter, S. M., Dawson, T. R. & Emeson, R. B. (1999) *Nature* **399**, 75-80.
27. Lonergan, K. M. & Gray, M. W. (1993) *Science* **259**, 812-6.
28. Price, D. H. & Gray, M. W. (1999) *Curr Genet* **35**, 23-9.
29. Price, D. H. & Gray, M. W. (1999) *RNA* **5**, 302-17.
30. Forget, L., Ustinova, J., Wang, Z., Huss, V. A. & Lang, B. F. (2002) *Mol Biol Evol* **19**, 310-9.
31. Laforest, M. J., Roewer, I. & Lang, B. F. (1997) *Nucleic Acids Res* **25**, 626-32.
32. Tomita, K., Ueda, T. & Watanabe, K. (1996) *Nucleic Acids Res* **24**, 4987-91.
33. Yokobori, S. & Pääbo, S. (1997) *J Mol Biol* **265**, 95-9.
34. Yokobori, S. & Pääbo, S. (1995) *Proc Natl Acad Sci U S A* **92**, 10432-5.
35. Yokobori, S. & Pääbo, S. (1995) *Nature* **377**, 490.
36. Reichert, A., Rothbauer, U. & Mörl, M. (1998) *J Biol Chem* **273**, 31977-84.
37. Börner, G. V., Mörl, M., Janke, A. & Pääbo, S. (1996) *EMBO J* **15**, 5949-57.
38. O'Kelly, C. J. (1993) *J Eukaryot Microbiol* **40**, 627-636.
39. Patterson, D. J. (1990) *J Mar Biol Ass U K* **70**, 381-393.
40. Flavin, M. & Nerad, T. A. (1993) *J Eukaryot Microbiol* **40**, 172-9.
41. Keiler, K. C., Shapiro, L. & Williams, K. P. (2000) *Proc Natl Acad Sci U S A* **97**, 7778-83.
42. Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M. & Gray, M. W. (1997) *Nature* **387**, 493-7.

43. Lang, B. F., Seif, E., Gray, M. W., O'Kelly, C. J. & Burger, G. (1999) *J Eukaryot Microbiol* **46**, 320-6.
44. Yang, Z. (1997) *Comput Appl Biosci* **13**, 555-6.
45. Gray, M. W. & Covello, P. S. (1993) *FASEB J* **7**, 64-71.
46. Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J. & Scott, J. (1987) *Cell* **50**, 831-40.
47. Zheng, H., Fu, T. B., Lazinski, D. and Taylor, J. (1992) *J Virol* **66**, 4693-4697.
48. Yokobori, S. I. & Pääbo, S. (1995) *Nature* **377**, 490.
49. Preiss, J., Dieckmann, M. & Berg, P. (1961) *J. Biol. Chem.* **236**, 1748-1757.
50. Burger, G., Plante, I., Lonergan, K. M. & Gray, M. W. (1995) *J Mol Biol* **245**, 522-37.
51. Lowe, T. M. & Eddy, S. R. (1997) *Nucleic Acids Res* **25**, 955-64.
52. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res* **22**, 4673-80.
53. Löytynoja, A. & Milinkovitch, M. C. (2001) *Bioinformatics* **17**, 573-4.
54. Hou, Y. M. & Schimmel, P. (1988) *Nature* **333**, 140-5.
55. Kunzmann, A., Brennicke, A. & Marchfelder, A. (1998) *Proc Natl Acad Sci U S A* **95**, 108-13.
56. Miller, J. P. & Philipps, G. R. (1971) *J Biol Chem* **246**, 1274-9.
57. Hong, Y., Cole, T. E., Brasier, C. M. & Buck, K. W. (1998) *Virology* **246**, 158-69.